Identification and Estimation of Nonseparable Single-Index Models in Panel Data with Correlated Random Effects

Cizek, P.; Lei, J.

Publication date:
2013

CentER

Discussion Paper

No. 2013-062

IDENTIFICATION AND ESTIMATION OF NONSEPARABLE SINGLE-INDEX MODELS IN PANEL DATA WITH CORRELATED RANDOM EFFECTS

By

Pavel Čížek, Jinghua Lei

7 November 2013

ISSN 0924-7815
ISSN 2213-9532

Tilburg University

Identification and estimation of nonseparable single-index models in panel data with correlated random effects

Pavel Čížek* Jinghua Lei†

November 7, 2013

Abstract

The identification of parameters in a nonseparable single-index models with correlated random effects is considered in the context of panel data with a fixed number of time periods. The identification assumption is based on the correlated random-effect structure: the distribution of individual effects depends on the explanatory variables only by means of their time-averages. Under this assumption, the parameters of interest are identified up to scale and could be estimated by an average derivative estimator based on the local polynomial smoothing. The rate of convergence and asymptotic distribution of the proposed estimator are derived along with a test whether pooled estimation using all available time periods is possible. Finally, a Monte Carlo study indicates that our estimator performs quite well in finite samples.

JEL codes: C14, C23

Key words: average derivative estimation, correlated random effects, local polynomial smoothing, nonlinear panel data

*CentER and Department of Econometrics and Operation Research, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands. Phone: +31-13-466-8723, Fax: +31-13-466-3280, E-mail: P.Cizek@tilburguniversity.edu

†Corresponding Author: Department of Econometrics and Operation Research, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands. E-mail: J.Lei@uvt.nl

1 Introduction

The single-index models, linking the response variable to regressors by means of a single linear combination, encompass a large number of practically applied models. To estimate these models, a significant amount of literature has been devoted in recent years to the local derivative and average derivative estimation. The average derivative estimation based on the Nadarava-Watson kernel regression (Gasser and Müller, 1984) was proposed and studied, for example, by Härdle and Stoker (1989) and Newey and Stoker (1993). As the local linear regression offers some advantages over the Nadaraya-Watson estimator (Fan and Gijbels, 1992), the average derivative estimation relying on the local polynomial regression was proposed by Hristache et al. (2001) and Li et al. (2003), for instance. Nevertheless, these classical estimators are primarily designed for cross-sectional data and the average derivative estimation for panel data is relatively scarce.

The main difficulty in dealing with nonlinear panel data is caused by the presence of individual specific heterogeneity, especially in the fixed effect models, which allow the individual effects to be correlated with the explanatory variables. Although the unobserved individual specific heterogeneity could be eliminated or treated as parameters to be estimated in linear or additive panel data models, such approaches cannot be readily applied to nonlinear panel data models as they results in inconsistent estimators due to the incidental parameters problem (Lancaster, 2000). Nevertheless, there has been a number of attempts to consistently estimate the nonlinear panel data models with specific model forms. For example, Manski (1987) and Charlier et al. (1995) proposed a (smoothed) maximum score estimator for discrete choice model; Honoré (1992) artificially censors the dependent variable in the censored regression model such that the individual fixed effect could be differenced away; Kyriazidou (1997) introduced a semiparametric method to estimate the parameters of sample selection models in panel data; and Abrevaya (1999) proposed a rank-based estimator for monotone transformation models. Additionally, there is also a branch of literature which aims at improving the performance of existing estimators that treat individual effects as parameters via bias-correction (e.g., see Hahn and Newey, 2004, or Bester and Hansen, 2009b). However, these approaches rely on parametric assumptions for specific structural model or on asymptotics, where both the number of observations and time dimension go to infinity.

Most recently, several papers have provided the identification and estimation for marginal effects in nonlinear panel models. Chernozhukov et al. (2013) derive bounds for marginal effects and propose two novel inference methods for parameters as solutions to nonlinear programs. Further, Bester and Hansen (2009a) achieve identification of average marginal effects in a correlated random effects (CRE) model by imposing that the individual-effect distribution depends on each covariate only through a scalar function of the values observed over time. Finally, the most similar work to the current paper is by Hoderlein and White (2012), who derive a generalized version of differencing that identifies local average responses in a general nonseparable model (without single-index structure). Considering two time periods and without assuming additional functional form restrictions or restrictions on the dependence between regressors and fixed effects, they identify effects for the subpopulation of individuals, who have not experienced a change in covariates between the two time periods.

Our identification strategy relies on an assumption similar to Bester and Hansen (2009a): the distribution of individual effects depends on the explanatory variables only by means of their time-averages. At the same time, the resulting estimator is close to Hoderlein and White (2012) in the sense that one estimates the derivatives of the first differences of a particular regression function. The crucial difference lies in the identification assumption of the CRE model, which is more restrictive assumptions than the one in Hoderlein and White (2012), but provides several practical advantages. First, our method can identify the regression coefficients and marginal effects for the whole population rather than for a subpopulation only. Second, although two time periods are also sufficient for identification, we do not restrict the estimation to only two time periods and make an explicit use of multiple time periods to improve estimation (this also renders a stability test if more than two time periods are available). Finally, let us mention that the model and its estimation – being based on a general nonlinear model – suits many applications such as those relying on various discrete-choice and limited-dependent-variables models as discussed in Hoderlein and White (2012) in details.

In the rest of the work, we first show how the parameters of interest are identified in Section 2. Next, a semiparametric average derivative estimation procedure is developed in Section 3, which is easy to compute and does not require numerical optimization. The rate of convergence and asymptotic distribution of the proposed estimator are derived in Section 4 and the finite sample performance of the procedure is documented by Monte Carlo experiments in Section 5. Proofs are included in the Appendices.

2 Identification

The panel data consist of n observations of time series $Y_i = (Y_{i1}, \ldots, Y_{iT})'$ and $X_i = (X_{i1}, \ldots, X_{iT})'$ for a dependent variable Y_{it} and a vector of explanatory variables X_{it}, which are independent and identically distributed across individual observations $i \in \{1, \ldots, n\}$. The number T of time periods is assumed to be finite and fixed. A general nonseparable model with an unobserved individual effects α_i can be described as

$$Y_{it} = \phi(X_{it}, \alpha_i, U_{it}),$$

where the individual effect α_i may be a vector of any finite dimension and U_{it} represents unobservables. In this paper, we assume more structure in that the explanatory variables enter into the mean response function only through a single linear index such that

$$Y_{it} = \phi(X_{it}'\beta, \alpha_i, U_{it}), \tag{1}$$

where β is a vector parameter that is common to all individuals i and α_i is a scalar or vector of individual fixed effects. This class of single-index models includes panel-data censored and truncated Tobit models (e.g., $Y_{it} = \max\{X_{it}'\beta + \alpha_i + U_{it}, 0\}$), binary choice models (e.g., $Y_{it} = I(X_{it}'\beta + \alpha_i + U_{it} > 0)$), or duration models with unobserved individual heterogeneity and random censoring. Our interest lies in the effect of X_{it} on Y_{it}, that is, we aim to estimate parameters β and the marginal effects of X_{it} on Y_{it}.

First, the assumptions for the identification of β are introduced.

Assumption 1. *Let (Ω, F, P) be a complete probability space on which are defined the random vectors $\alpha_i : \Omega \to \mathcal{A}, \mathcal{A} \subseteq \mathbb{R}^{d_a}$, and $(Y_{it}, X_{it}, U_{it}) : \Omega \to \mathcal{Y} \times \mathcal{X} \times \mathcal{U}$, $\mathcal{Y} \subseteq \mathbb{R}, \mathcal{X} \subseteq \mathbb{R}^d, \mathcal{U} \subseteq \mathbb{R}^{d_u}$, for any $i \in \mathbb{N}$, $t = 1, \ldots, T$, and finite integers d_a, d, d_u, and T. Let for all $i \in \mathbb{N}$ and $t = 1, \ldots, T$ hold that: (i) $E(|Y_{it}|) < \infty$; (ii) $Y_{it} = \phi(X_{it}'\beta, \alpha_i, U_{it})$, where*

$\beta \in \mathbb{R}^d$ is a vector of d parameters and ϕ is an unknown Borel-measurable function, which is not constant on the support of $X_{it}'\beta$ for any $(\alpha_i, U_{it}) \in \mathcal{A} \times \mathcal{U}$; and (iii) realizations of (Y_{it}, X_{it}) are observable, whereas those of (α_i, U_{it}) are not observable.

Assumption 1 formally specifies the data generating process and is similar to Assumption A1 of Hoderlein and White (2012). While we allow for more than two time periods, we impose a functional form restriction: the exact functional form may be unknown, but the dependent variable Y_{it} depends on the explanatory variables X_{it} only by means of a linear index $X_{it}'\beta$. A general data generating process without a single-index structure will be discussed later in the case when a researcher is interested only in the average partial effects of X_{it} on Y_{it}. Further, α_i in the above model is unobserved and time invariant and can be correlated with the covariates X_{it} (see Assumption 3 below).

Assumption 2. *Unobservables U_{it} are independent of α_i and X_{is} and identically distributed for all $i = 1, \ldots, n$ and $s, t = 1, \ldots, T$.*

Assumption 2 is the strict exogeneity assumption. The idiosyncratic error term U_{it} is assumed to be uncorrelated with the explanatory variables of all past, current, and future time periods of the same individual. Although this is stronger than Assumption A3 of Hoderlein and White (2012), dependence between the 'usual' error term and the explanatory variables is not ruled out by Assumption 2. For example, a linear panel data model $Y_{it} = \alpha_i + X_{it}'\beta + \varepsilon_{it}$, where $\varepsilon_{it} = g(\alpha_i, X_{it}'\beta)U_{it}$, satisfies Assumptions 1–2, but exhibits heteroscedasticity depending on the linear index and the individual effect. Assumption 2 however rules out the presence of lagged dependent variables: the weakest form of Assumption 2 required here is that U_{it} is independent of $(X_{i(t-1)}, X_t, X_{i(t+1)})$. The model thus cannot possess dynamics.

The next assumption formulates the main identification restriction on the explanatory variables that are related to the individual effects α_i.

Assumption 3. *Let us assume that (i) there are no time-constant covariates in X_{it}, that (ii) random vectors X_{it} are identically and continuously distributed for all $i \in \mathbb{N}$ and $t = 1, \ldots, T$, and that, for some fixed $1 \leq t' < T$, (iii) the joint distributions $F_{X_t, X_{t-t'}}$ of $(X_{it}, X_{i(t-t')})$ are identical for all $i \in \mathbb{N}$ and $t' < t \leq T$. Then the conditional distribution $F_{\alpha|X_t, X_{t-t'}}$ of the individual effect is assumed to depend on X_t and $X_{t-t'}$ only by means of* *their average:* $F_{\alpha|X_t,X_{t-t'}}(\alpha_i|X_{it},X_{i(t-t')}) = F_{\alpha|X_t,X_{t-t'}}(\alpha_i|X_{it}+X_{i(t-t')})$*. Additionally, the defined distribution functions are twice continuously differentiable with uniformly bounded derivatives on* $\mathcal{X}$*.*

Assumption 3 is the key assumption for the identification of β. Similarly to other estimation methods that rely on some kind of differencing across time, variables constant over time cannot be included in the model. More importantly, Assumption 3 restricts the process $\{X_{it}\}$ and the joint process $\{X_{it}, X_{i(t-t')}\}$ to be identically distributed for a fixed time gap t' (the time gap t', $1 \leq t' < T$, is a fixed quantity from now on unless stated otherwise). In particular, the joint distribution $F_{X_t,X_{t-t'}}(X_{it}, X_{i(t-t')})$ is assumed to be time invariant in order to estimate using jointly all available time periods (this will be further discussed and tested by means of a χ^2 statistics in Section 4). This assumption is however not necessary for estimation at any fixed time t. Finally, while α_i and X_{it} or $X_{i(t-t')}$ depend on each other in general, the CRE models impose, for instance, that the individual effects α_i depend on covariates X_{it} only via their time-averages $\bar{X}_i = T^{-1}\sum_{i=1}^{T} X_{it}$ (see Bester and Hansen, 2009a, for a discussion of various CRE assumptions). In the case of two time periods, this implies in Assumption 3 that the conditional distribution $F_{\alpha|X_t,X_{t-t'}}(\alpha_i|X_{it}, X_{i(t-t')})$ depends only on the sum $X_{it} + X_{i(t-t')}$ rather than individual values. (Bester and Hansen, 2009a, argued that it is not possible to identify marginal effect using two time periods in the general CRE model if $F_{\alpha|X_t,X_{t-t'}}(\alpha_i|X_{it}, X_{i(t-t')}) = F_{\alpha|X_t,X_{t-t'}}(\alpha_i|h(X_{it}, X_{i(t-t')}))$ with a general unknown function h).

To derive the main identification result, additional regularity assumptions are needed: differentiability of the function ϕ and existence of an integrable majorant to enable the interchange of integration and differentiation. The abbreviated notation $F(\alpha|x_t, x_{t-t'}) \equiv F_{\alpha|X_t,X_{t-t'}}(\alpha|X_{it} = x_t, X_{i(t-t')} = x_{t-t'})$ and $f(\alpha|x_t, x_{t-t'}) \equiv f_{\alpha|X_t,X_{t-t'}}(\alpha|X_{it} = x_t, X_{i(t-t')} = x_{t-t'})$ is used.

Assumption 4. *The function* $\phi(v,\alpha,u)$ *is twice continuously differentiable with respect to* $v \in \mathbb{R}$ *for each* $(\alpha, u) \in \mathcal{A} \times \mathcal{U}$*. Moreover,* $E[\phi'_{xb}(X'_{it}\beta, \alpha_i, U_{it})] < \infty$*, where* $\phi'_{xb}(v,\alpha,u) = \partial\phi(v,\alpha,u)/\partial v$*.*

Assumption 5. *For each* $(x_t, x_{t-t'}) \in \mathbb{R}^d \times \mathbb{R}^d$*, there exists a* σ*-finite measure* $\mu(.|x_t, x_{t-t'})$ *absolutely continuous with respect to* $F(\cdot|x_t, x_{t-t'})$ *so that there exists a Radon-Nikodym* *density f such that $F(d\alpha|x_t, x_{t-t'}) = f(\alpha|x_t, x_{t-t'})\mu(d\alpha|x_t, x_{t-t'})$ for each $\alpha \in \mathcal{A}$.*

Assumption 6. *For each $(x_t, x_{t-t'}) \in \mathbb{R}^d \times \mathbb{R}^d$ and $t' < t \leq T$, there exists an integrable dominating function $D(\alpha_i, U_{it}|x_t, x_{t-t'})$ such that*

$$\sup_{\|v-x'\beta\|<\epsilon} |\phi'_{xb}(v, \alpha_i, U_{it}) f(\alpha_i|x_t, x_{t-t'})| \leq D(\alpha_i, U_{it}|x_t, x_{t-t'}),$$

$$\sup_{\|v-x'\beta\|<\epsilon} \left| \phi(v, \alpha_i, U_{it}) \frac{\partial f(\alpha_i|x_t, x_{t-t'})}{\partial x} \right| \leq D(\alpha_i, U_{it}|x_t, x_{t-t'}).$$

The main identification result is presented below under Assumptions 1–6. Although we assume for simplicity that the explanatory variables are continuously distributed, discrete regressors can be also included under the assumptions analogous to other identification results for single-index models (e.g., Ichimura, 1993). Furthermore, note that the result of the following theorem holds both if the expectations are taken across all cross-sectional units and all time periods as well as if the expectations are taken only across cross-sectional units for a fixed time period $t, t' < t \leq T$.

Theorem 1. *Under the Assumption 1-6, β is identified up to scale by*

$$\beta = \left\{ E\left(E\left[\phi'_{xb}(X'_{it}\beta, \alpha_i, U_{it}) | X_{it}, X_{i(t-t')} \right] \right) \right\}^{-1} \times E\left\{ \frac{\partial}{\partial X_{it}} E[(Y_{it} - Y_{i(t-t')})|X_{it}, X_{i(t-t')}] \right\}$$
$$\propto E\left\{ \frac{\partial}{\partial X_{it}} E[(Y_{it} - Y_{i(t-t')})|X_{it}, X_{i(t-t')}] \right\},$$

provided that the denominator is finite and non-zero. Moreover, when $\phi(X'_{it}\beta, \alpha_i, U_{it}) = X'_{it}\beta + \psi(\alpha_i, U_{it})$, β is point identified.

Theorem 1 states that the parameters of interest are propotional to quantity $\delta_{t'} = E\{\partial E[(Y_{it} - Y_{i(t-t')})|X_{it}, X_{i(t-t')}]/\partial X_{it}\}$, that is, to the average derivative of $E[Y_{it} - Y_{i(t-t')}|X_{it}, X_{i(t-t')}]$ with respect to X_{it}. By estimating $\delta_{t'}$, β will thus be estimated up to scale. The identification of β requires only two time periods when considering to the simplest case $t' = 1$; the estimator based on $t' = 1$ is further referred to as the first-difference average derivative estimation $\delta_{FD} = \delta_1$.

Remark 1. In Theorem 1, t' can equal to any integer smaller than the total number T of time periods. Although we primarily concentrate on the case of one fixed t' here, one can actually obtain $T-1$ different estimators based on $\delta_1, \delta_2, \ldots, \delta_{T-1}$ for any given T. In general, $\delta_{t'} \neq \delta_{t''}$ for $t' \neq t''$ because we have not imposed that the joint distribution $(X_{it}, X_{i(t-t')})$ is the same as that of $(X_{it}, X_{i(t-t'')})$. To solve the problem of different average derivatives and estimators for different values of t', they can be normalized by their norms $||\delta_{t'}||$ – using $\delta_{t'}/||\delta_{t'}|| = \beta/\|\beta\|$ for all $t' < T$ – since the parameters are identified only up to scale. Subsequently, it is possible to combine their information and to base the weighted average derivative estimator on

$$\delta_W = \sum_{t'=1}^{T-1} w_{t'} \cdot \delta_{t'}/||\delta_{t'}||, \tag{2}$$

where $w_{t'}$ represents suitably chosen weights (e.g., proportional to the variance of $\delta_{t'}$). Finally, note that when $\phi(X'_{it}\beta, \alpha_i, U_{it}) = X'_{it}\beta + \psi(\alpha_i, U_{it})$, that is, in the linear panel models with non-additive errors, Theorem 1 shows that β is point identified and there is no need for scale normalization. In this case, $\delta_{t'} = \delta_{t''}$ for any $t' \neq t''$, which renders other possibilities how to combine estimates $\delta_1, \delta_2, \ldots, \delta_{T-1}$; see Section 3 for more details.

Remark 2. As in Bester and Hansen (2009a), if the interest of researcher lies only in the partial effects of X_{it} on Y_{it} with individual heterogeneity held constant, that is, in the partial effects averaged over the distribution of individual-specific effects, then our model specification with its single-index structure could be relaxed to a general nonseparable structure $Y_{it} = \phi(X_{it}, \alpha_i, U_{it})$. If the average partial effect is defined by $\delta_{APE} = E[\partial\phi(X_{it}, \alpha_i, U_{it})/\partial X_{it}]$, then this average partial effects δ_{APE} is identified and could also be estimated by

$$\delta_{APE} = E\left[\frac{\partial}{\partial X_{it}} E[(Y_{it} - Y_{i(t-t')})|X_{it}, X_{i(t-t')}]\right]$$

The proof is analogous to the proof of Theorem 1 and is omitted here.

3 Estimation

To estimate the expectation $\delta_{t'} = E\{\partial E[Y_{it} - Y_{i(t-t')}|X_{it}, X_{i(t-t')}]/\partial X_t\}$ for given $t', 1 \leq t' < T$, we first estimate the expectation $E[Y_{it} - Y_{i(t-t')}|X_{it}, X_{i(t-t')}]$ and its derivative by means of the local linear or polynomial regression and then the outer expectation will be evaluated. Later, we derive the asymptotic distribution of the proposed estimator $\widehat{\delta}_{t'}$ of $\delta_{t'}$, relying on the properties of the local polynomial estimators, including the uniform strong consistency and asymptotic normality, established by Masry (1996) and Masry (1997) for general mixing processes.

Let us denote $\Delta Y_{it,t'} = Y_{it} - Y_{i(t-t')}$ and $Z_{it,t'} = (X'_{it}, X'_{i(t-t')})'$. For the local polynomial regression, non-negative kernel weights $K_h(t) = K(t/h_n)/h_n^d$ are used, where the bandwidth h_n is for simplicity common to all variables. To estimate the expectation $m(z) = E[\Delta Y_{it,t'}|Z_{it,t'} = z]$ and its derivatives $\delta_{t'}(z) = m'_1(z) = \partial E[(Y_{it} - Y_{i(t-t')})|Z_{it,t'} = z]/\partial X_{it}$ and $\tilde{\delta}_{t'}(z) = m'_2(z) = \partial E[(Y_{it} - Y_{i(t-t')})|Z_{it,t'} = z]/\partial X_{i(t-t')}$, we can consider the local linear regression minimizing

$$\sum_{i=1}^{n}\sum_{t=t'+1}^{T}\left[\Delta Y_{it,t'} - b_{0,t'}(z) - (Z_{it,t'} - z)b_{1,t'}(z)\right]^2 K_h\left(Z_{it,t'} - z\right), \tag{3}$$

where the least squares estimate $\widehat{b}_{t'}(z) = (\widehat{b}_{0,t'}(z), \widehat{b}'_{1,t'}(z))'$ estimates (i) $m(z)$ by the only element of $\widehat{b}_{0,t'}(z)$ and (ii) the derivatives $\partial m(z)/\partial z = (\delta_{t'}(z)', \tilde{\delta}_{t'}(z)')'$ by the $2d$ elements of $\widehat{b}_{1,t'}(z)$. Similarly to Härdle and Stoker (1989), the local linear estimator would require that a kernel K of order $p > 2d$ is used to guarantee the $\sqrt{n}$ consistency of the average derivative estimator proposed later. An alternative lies in the use of the local polynomial regression of order $p > d$, which includes higher powers of $Z_{it,t'} - z$ in (3). Denoting $|k| = k_1 + \ldots + k_{2d}$ the "length" of a vector $k = (k_1, \ldots, k_{2d}) \in \mathbb{N}_0^{2d}$ and understanding $z^k = z_1^{k_1} \times \ldots \times z_{2d}^{k_{2d}}$, the local polynomial estimator can be defined as a minimizer of

$$\sum_{i=1}^{n}\sum_{t=t'+1}^{T}\left[\Delta Y_{it,t'} - \sum_{|k|=0}^{p}(Z_{it,t'} - z)^k b_{k,t'}(z)\right]^2 K_h\left(Z_{it,t'} - z\right); \tag{4}$$

the parameters of interest are again the $2d$ elements of $\widehat{b}_{1,t'}(z)$, which estimate $\partial m(z)/\partial z = (\delta_{t'}(z)', \tilde{\delta}_{t'}(z)')'$.

The least squares solution of (3) can be explicitly formulated using some matrix notation. Denoting the vectors of responses $\Delta Y_{i,t'} = (\Delta Y_{i(t'+1),t'}, \ldots, \Delta Y_{iT,t'})'$ and $\Delta Y_{t'} = (\Delta Y'_{1,t'}, \ldots, \Delta Y'_{n,t'})'$, the corresponding matrices of the explanatory variables $Z_{i,t'}(z) = (Z_{i(t'+1),t'} - z, \ldots, Z_{iT,t'} - z)'$ and $Z_{t'}(z) = \{(Z^k_{i,t'}(z))^p_{|k|=0}\}^n_{i=1}$ (where the latter includes intercept), and the matrix of kernel weights $W_{t'}(z) = \mathrm{diag}\{K_h(Z_{it,t'} - z)\}_{i=1,t=t'+1}^{n,\ T}$, the estimate minimizing (4) equals

$$\widehat{b}_{t'}(z) = (\widehat{b}'_{0,t'}(z), \widehat{b}'_{1,t'}(z), \ldots, \widehat{b}'_{p,t'}(z))' = [Z_{t'}(z)'W_{t'}(z)Z_{t'}(z)]^{-1}Z_{t'}(z)'W_{t'}(z)\Delta Y_{t'}$$

(provided that $Z_{t'}(z)$ has a full rank, see also Assumption 3).

Given that we are interested in estimating $\delta_{t'}(z)$, that is, the first to dth elements of $b_{1,t'}(z)$, the local derivative estimator of $\delta_{t'}(z)$ is given by

$$\widehat{\delta}_{t'}(z) = L\widehat{b}_{t'}(z) = L \cdot [Z_{t'}(z)'W_{t'}(z)Z_{t'}(z)]^{-1}Z_{t'}(z)'W_{t'}(z)\Delta Y_{t'}, \tag{5}$$

where $L = (e_2, \ldots, e_{d+1})'$ and e_j represents a vector with its jth element equal to 1 and all other elements equal to 0. Note that in (5) and in the case of other estimators, the dependence on the size n of cross-section units is marked by the hat and is kept implicit to avoid clutter (the asymptotic properties will be derived for $n \to \infty$, while T is fixed).

Recalling Theorem 1, the parameters β are proportional to $\delta_{t'} = E[\delta_{t'}(z)]$. The finite-sample average derivative estimator of β can thus be defined for a given t' as

$$\widehat{\delta}_{t'} = \frac{1}{n(T-t')}\sum_{i=1}^{n}\sum_{t=t'+1}^{T}\widehat{\delta}_{t'}(Z_{it,t'}) = \frac{1}{n(T-t')}\sum_{i=1}^{n}\sum_{t=t'+1}^{T} L \cdot \widehat{b}_{t'}(Z_{it,t'}). \tag{6}$$

Since β is identified only up to scale, this estimator in equation (6) should be scale normalized.

Remark 3. The weighted average derivative estimator corresponding to (2) can be defined by

$$\widehat{\delta}_W = \sum_{t'=1}^{T-1} w_{nt'} \cdot \widehat{\delta}_{t'}/||\widehat{\delta}_{t'}||, \tag{7}$$

where weights $w_{nt'}$ can possibly depend on the sample size.

Remark 4. For the linear models with non-additive errors

$$\phi(X'_{it}\beta, \alpha_i, U_{it}) = X'_{it}\beta + \psi(\alpha_i, U_{it}), \tag{8}$$

Theorem 1 indicates that β is point identified. In this case, the pooling of estimators for different levels of t' can be replaced by pooling the objective functions as the identified parameters $\delta_1, \ldots, \delta_{T-1}$ are equal to β for all $1 \leq t' < T$. In particular, the pairwise-difference local derivative estimator $\hat{\delta}_{PW}(z)$ can be defined as $L\widehat{b}_{(*)}(z)$, where $\widehat{b}_{(*)}(z) = (\widehat{b}'_{0,(*)}(z), \ldots, \widehat{b}'_{p,(*)}(z))'$ minimizes, for example, in the case of the local polynomial regression (4),

$$\sum_{t'=1}^{T-1} \sum_{i=1}^{n} \sum_{t=t'+1}^{T} \left[\Delta Y_{it,t'} - \sum_{|k|=0}^{p} (Z_{it,t'} - z)^k b_{|k|,(*)}(z) \right]^2 K_h(Z_{it,t'} - z);$$

that is, the minimization is performed jointly across all values of t'. The corresponding pairwise-difference average derivative estimator of β then equals

$$\widehat{\delta}_{PW} = \frac{2}{nT(T-1)} \sum_{t'=1}^{T-1} \sum_{i=1}^{n} \sum_{t=t'+1}^{T} \hat{\delta}_{PW}(Z_{it,t'}) = \frac{2}{nT(T-1)} \sum_{t'=1}^{T-1} \sum_{i=1}^{n} \sum_{t=t'+1}^{T} L\widehat{b}_{(*)}(Z_{it,t'}). \quad (9)$$

Remark 5. Note that this pairwise-difference average derivative estimator is also applicable in nonlinear random-effect models. Contrary to the fixed-effect or CRE models, where the multiplicative constants in Theorem 1, $E\{E[\phi'_{xb}(X'_{it}\beta, \alpha_i, U_{it})|X_{it}, X_{i(t-t')}]\}$, are generally different for various t', the random-effect specification, where the individual effects are independent of the covariates X_t, implies that $E[\phi'_{xb}(X'_{it}\beta, \alpha_i, U_{it})|X_{it}, X_{i(t-t')}] = E[\phi'_{xb}(X'_{it}\beta, \alpha_i, U_{it})|X_{it}]$ does not depend on t'. Thus, the scaling coefficients in Theorem 1 are independent of t' as in the case of the linear model (8). The parameters β are however estimated only up to scale in the random-effect model.

The proposed estimator $\widehat{\delta}_{t'}$ in equation (6) is similar to the least-square average derivative estimator of Li et al. (2003), but the underlying data are no longer independent and identically distributed in our case (e.g., because of the individual effects α_i). As the number T of time periods is finite, the dependence is however limited to a fixed number of time periods. To establish the uniform consistency of the local derivative estimator $\widehat{\delta}_{t'}(z)$ and the consistency and asymptotic distribution of the average derivative estimator $\widehat{\delta}_{t'}$ based on the local polynomial regression (4), the following assumptions are used (in the case of the estimator based on the local linear smoothing (3), the kernel function would instead be of order p).

Assumption 7.

1. *As $n \to +\infty$, the bandwidth h_n satisfies $nh_n^{2p+2} \to 0$ and $nh_n^{2d+2}/\ln n \to \infty$.*

2. *The kernel K is bounded with a compact support and $\int K(u)du = 1$, $\int uK(u)du = 0$, and $\int uu'K(u)du = cI_{2d}$ for some $c > 0$, where I_k denotes the $k \times k$ identity matrix.*

3. *Let $\mathcal{D} \subset \mathbb{R}^{2d}$ denote a compact support of the identically distributed random vectors $Z_{it,t'} = (X'_{it}, X'_{i(t-t')})'$ and assume that the density function f_z of $Z_{it,t'}$ exists, is bounded, and twice continuously differentiable.*

4. *Further, let function $m(z) = E[\Delta Y_{it,t'}|Z_{it,t'} = z]$ be $(p+1)$ times differentiable with all partial derivatives being uniformly bounded and Lipschitz on $\mathcal{D} \subset \mathbb{R}^{2d}$ and let $m(Z_{it,t'})$ and its $(p+1)$ derivatives in absolute values have finite expectations.*

5. *Finally, errors $V_{it,t'} = E(\Delta Y_{it,t'}|Z_{it,t'}) = \Delta Y_{it,t'} - m(Z_{it,t'})$ have finite fourth moments. Assume that (co)variances $\sigma^2_{t'}(z) = E(V^2_{it,t'}|Z_{it,t'} = z)$ and $\sigma_{ts,t'}(z_1, z_2) = E(V_{it,t'}V_{is,t'}|Z_{it,t'} = z_1, Z_{is,t'} = z_2)$ for $t' < s \leq T$ and $t' < t \leq T$ are continuous in z and (z_1, z_2), respectively.*

Assumptions 7.1 and 7.2 are standard assumptions on the bandwidth and kernel in the average derivative estimation (e.g., Härdle and Stoker, 1989, and Li et al., 2003). Additionally, Assumption 7.3 imposes that the explanatory variables have a compact support. If this common assumption in the semiparametric literature is not satisfied, it can be imposed by means of trimming (see Li and Racine, 2007, Chapter 8, for various examples). The existence of $p+1$ derivatives in Assumption 7.4, which is also reflected implicitly in Assumption 7.1 as $p > d$, is also common to many average derivative estimators (e.g., Härdle and Stoker, 1989). If inconvenient, it can be relaxed by estimating iteratively using the procedure of Hristache et al. (2001), which requires only the existence of the second derivatives irrespective of the dimension d.

Under the stated assumptions, the uniform consistency of $\widehat{\delta}_{t'}(z)$ follows directly from Masry (1996, Theorem 4).

Theorem 2. *Under Assumptions 1–7, $\widehat{\delta}_{t'}(z)$ is uniformly consistent on $\mathcal{D}$:* $\sup_{z\in D}|\widehat{\delta}_{t'}(z)-\delta_{t'}(z)| = O(\ln n \cdot [n^{-1/2}h_n^{-d-1}]) + O(h_n^p)$ *as $n \to +\infty$.*

Using the consistency of $\widehat{\delta}_{t'}(z)$, the asymptotic distribution of the average derivative estimator can be now derived.

Theorem 3. *Under Assumptions 1–7, the average derivative estimator $\widehat{\delta}_{t'}$ defined in equation* (6) *for some $t', 1 \le t' < T$, is consistent and asymptotically normal:*

$$\sqrt{n}\left(\widehat{\delta}_{t'} - E[m_1'(Z_{it,t'})] - h_n^p L A_{t'}\right) \xrightarrow{F} N\left(0, \Phi_{t'} + \Omega_{t'}\right),$$

where $A_{t'}$ is defined in Lemma 1,

$$\Phi_{t'} = \frac{1}{(T-t')^2}\sum_{t=t'+1}^{T}\sum_{s=t'+1}^{T} E\left[\sigma_{st,t'}(Z_{it,t'}, Z_{is,t'})\cdot G_{[d],1}(Z_{it,t'})G_{[d],1}(Z_{is,t'})'\right]$$

with $G_{[d],1}(z) = L[M^f(z)]^{-1}Q^f(z)e_1$ and matrices of kernel weights $M^f(z)$ and $Q^f(z)$ are defined in Appendix B, and

$$\Omega_{t'} = \frac{1}{(T-t')^2}\sum_{t=t'+1}^{T}\sum_{s=t'+1}^{T} Cov\left[m_1'(Z_{it,t'}), m_1'(Z_{is,t'})\right].$$

Theorem 3 proves that the average derivative estimator $\widehat{\delta}_{t'}$ of $E[m_1'(Z_{it,t'})]$, which equals β up to scale, $\beta \propto E[m_1'(Z_{it,t'})]$, is consistent and asymptotically normal for any given t'. The bias term $h_n^p L A_{t'}$ is generally present as we assume only $\sqrt{n}h_n^{p+1} \to 0$, where p denotes the order of the local polynomial approximation. It becomes negligible if $\sqrt{n}h_n^p \to 0$ by choosing a large order of the polynomial or a smaller bandwidth. The asymptotic variance of the estimator resembles the result of Li et al. (2003) as it consists of two components corresponding to the asymptotic variance of $\sqrt{n}\{\widehat{\delta}_{t'} - h_n^p L A_{t'} - n^{-1}\sum_{i=1}^n m_1'(Z_{it,t'})\}$ and $\sqrt{n}\{n^{-1}\sum_{i=1}^n m_1'(Z_{it,t'}) - E[m_1'(Z_{it,t'})]\}$, respectively, which are asymptotically independent. The asymptotic variance however does not depend here only on the expected variance of the errors and first-order conditions, but also on their covariances over time as the regression errors $V_{it,t'} = \Delta Y_{it,t'} - m(Z_{it,t'})$ can exhibit heteroscedasticity and serial correlation.

4 Test of stationarity

In Sections 2 and 3, a weak form of stationarity of X_{it}, requiring the pairs of data from periods t and $t-t'$ to be jointly identically distributed across time t, is assumed so that estimation of β by $\delta_{t'} = E\{\partial E[(Y_{it} - Y_{i(t-t')})|X_{it}, X_{i(t-t')}]/\partial X_{it}\}$ can be based on all available time periods. In this section, we focus on constructing a test of this assumption, provided that one observes the data for at least three time periods. The estimation procedure actually relies on the implication of the stationarity assumption that the expectation $\delta_{s,t'} = E\{\partial E[(Y_{it} - Y_{i(t-t')})|X_{it}, X_{i(t-t')}]/\partial X_{it}|t = s\}$ evaluated at one fixed time $s, t' < s \leq T$, does not depend on the time point s. Thus, this implication can be stated as $\delta_{s,t'} = \delta_{s',t'}$ for all s and s', $t' < s, s' \leq T$, that is, different pairs of time periods $(s, s-t')$ and $(s', s'-t')$ produce the same estimates (the time difference t' is fixed).

If the number T of time periods is larger than $t'+1$ (e.g., $T > 2$ if $t' = 1$), there are $T - t' - 1$ possible expressions for $\delta_{t'}$: $\delta_{t'+1,t'}, \ldots, \delta_{T,t'}$, which are all equal under Assumptions 1–6, see Theorem 1. Denoting all these expressions as $\delta^*_{t'} = (\delta'_{t'+1,t'}, \ldots, \delta'_{T,t'})'$, we will thus test the null hypothesis that $\delta_{t'+1,t'} = \delta_{t'+2,t'} = \cdots = \delta_{T,t'}$:

$$\begin{aligned} H_0 &: R_{t'}\delta^*_{t'} = 0 \\ H_a &: R_{t'}\delta^*_{t'} \neq 0, \end{aligned} \tag{10}$$

where the $[d(T-t'-1)] \times [d(T-t')]$ matrix $R_{t'}$ can be expressed (using ι_k as a symbol for the vector with length k and all elements equal to 1) as

$$R_{t'} = \begin{pmatrix} I_d & -I_d & 0 & \ldots & 0 \\ I_d & 0 & -I_d & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ I_d & 0 & 0 & \ldots & -I_d \end{pmatrix} = \left[\iota_{(T-t'-1)}, -I_{(T-t'-1)\times(T-t'-1)}\right] \otimes I_d.$$

As Theorem 3 also establishes the asymptotic distribution of $\sqrt{n}(\widehat{\delta}_{s,t'} - h_n^p L A_{t'} - E[m'_1(Z_{is,t'})])$ for any given $t' < s \leq T$ and $h_n^p L A_{t'} - E[m'_1(Z_{is,t'})]$ have the same value under the null hypothesis for all s, the test (10) is equivalent to

$$H_0 : R_{t'}\bar{\delta}^*_{t'} = 0$$
$$H_a : R_{t'}\bar{\delta}^*_{t'} \neq 0,$$

where $\bar{\delta}^*_{t'} = (\bar{\delta}'_{t'+1,t'}, \ldots, \bar{\delta}'_{T,t'})'$ and $\bar{\delta}_{s,t'} = (\widehat{\delta}_{s,t'} - h^p LA_{t'} - E[m'_1(Z_{is,t'})])$ for all $s = t'+1, \ldots, T$.

Let us now establish the asymptotic distribution of $\bar{\delta}^*_{t'}$.

Theorem 4. *Under Assumptions 1–7 and the null hypothesis $R_{t'}\bar{\delta}^*_{t'} = 0$, we have*

$$\sqrt{n}\bar{\delta}^*_{t'} \xrightarrow{F} N\left(0, \widetilde{\Phi}_{t'} + \widetilde{\Omega}_{t'}\right),$$

where $\widetilde{\Phi}_{t'}$ and $\widetilde{\Omega}_{t'}$ are square matrices consisting of $(T-t'-1)\times(T-t'-1)$ blocks of dimensions $d \times d$; the blocks with coordinates (k,l) within matrices $\widetilde{\Phi}_{t'}$ and $\widetilde{\Omega}_{t'}$ have the forms $\widetilde{\Phi}^{(k,l)}_{t'} = E[\sigma_{kl,t'}(Z_{ik,t'}, Z_{il,t'}) \cdot G_{[d],1}(Z_{it,t'})G_{[d],1}(Z_{is,t'})']$ and $\widetilde{\Omega}^{(k,l)}_{t'} = Cov[m'_1(Z_{ik,t'}), m'_1(Z_{il,t'})]$, where $G_{[d],1}(z) = L[M^f(z)]^{-1}Q^f(z)e_1$ and matrices of kernel weights $M^f(z)$ and $Q^f(z)$ are defined in Appendix B.

Theorem 4 implies that a χ^2 test statistics can be constructed in the following way: since the (asymptotic) distributions of $R_{t'}\bar{\delta}^*_{t'}$ and $R_{t'}\delta^*_{t'}$ are identical under H_0, it holds that

$$TS_{t'} = n(R_{t'}\delta^*_{t'})'\left(\widetilde{\Phi}_{t'} + \widetilde{\Omega}_{t'}\right)^{-1}(R_{t'}\delta^*_{t'}) \xrightarrow{F} \chi^2\{d\left(T-t'-1\right)\}$$

under H_0. Using a consistent estimate $\widehat{\widetilde{\Phi}}_{t'} + \widehat{\widetilde{\Omega}}_{t'}$ of $\widetilde{\Phi}_{t'} + \widetilde{\Omega}_{t'}$ and the average derivative estimator $\hat{\delta}^*_{t'}$ of $\delta^*_{t'}$, it follows that

$$\widehat{TS}_{t'} = n(R_{t'}\hat{\delta}^*_{t'})'\left(\widehat{\widetilde{\Phi}}_{t'} + \widehat{\widetilde{\Omega}}_{t'}\right)^{-1}(R_{t'}\hat{\delta}^*_{t'}) \xrightarrow{F} \chi^2\{d\left(T-t'-1\right)\}.$$

The null hypothesis H_0 is thus rejected against the alternative H_a at significance level α if $\widehat{TS}_{t'} > \chi^2_\alpha\{d\,(T-t'-1)\}$. In such a case, the estimator (6) based on all observed time periods cannot be used, but instead a weighted average of estimators obtained at various time periods t has to be employed similarly to (7).

5 Simulation results

This section reports evidence on the finite sample behavior of estimators constructed using the proposed identification principle for several classical panel models with correlated random effects. The aim is to compare the average-derivative estimator with the existing procedures designed specifically for each individual model. The data generating processes exhibit two important features: nonzero correlation between individual effects and the covariates and strictly stationary covariates.

The models contain two stationary regressors X_{1it} and X_{2it}, which are independent for all i and generated by $X_{1it} = X_{1it-1}/2 + u_{1it}$, $\quad X_{1i0} \sim N(0, 4/3)$, $\quad X_{2it} = X_{2it-1}/3 + u_{2it}$, $\quad X_{2i0} \sim N(0, 9/8)$, $\quad u_{1it} \sim N(0, 1)$, $\quad u_{2it} \sim N(0, 1)$, where the number of individuals is $n = 100$ and the number of time periods is $T = 8$. The individual effects, generated by $\alpha_i = T^{-1} \sum_{t=1}^{T} X_{2it} + \eta_i$, $\quad \eta_i \sim U(-1/2, 1/2)$, are correlated with X_{2it}, but they are independent of X_{1it}. The true regression parameters are always $\beta_1 = 1$ and $\beta_2 = -1$.

We first consider the CRE linear model: $Y_{it} = X_{1it}\beta_1 + X_{2it}\beta_2 + \alpha_i + \epsilon_{it}$, where the ϵ_{it}'s are independently drawn from $N(0, 1)$. Next, we consider the binary-choice logistic model: $Y_{it} = I(X_{1it}\beta_1 + X_{2it}\beta_2 + \alpha_i + \epsilon_{it} > 0)$, where the ϵ_{it}'s are independently drawn from the standard logistic distribution. This models is analyzed in the case of homoscedastic logit with $\epsilon_{it} \sim \Lambda(0, 1)$, heteroscedastic logit with $\epsilon_{it} \sim \Lambda(0, \exp(1 + \alpha_i^2))$, and random-effect logit with $\epsilon_{it} \sim \Lambda(0, 1)$, where the individual effect $\alpha_i \sim N(0, 1)$. Finally, we consider the censored regression model: $Y_{it} = \max\{0, X_{1it}\beta_1 + X_{2it}\beta_2 + \alpha_i + \epsilon_{it}\}$, where the ϵ_{it}'s are independently drawn from $N(0, 1)$. For each model, 1000 samples are generated.

As there are two regressors in the original models, our multivariate local polynomial regression contains four regressors. The order of polynomials that we use in simulations is thus $p = 3$. To implement the average derivative estimators, choices need to be made for the kernel function K and the bandwidth h_n. We use the quartic kernel, noting that the choice of the kernel function has typically a rather limited impact on performance of nonparametric estimators. The bandwidth, which equals $\sigma_j h_n$ for each variable (σ_j denotes here the standard deviation of the jth covariate), is selected by the least-squares cross-validation method. The leave-one-out local polynomial estimator of $m_{t'}(Z_{it,t'})$ is obtained by $\widehat{m}^h_{-i,t'}(Z_{it,t'}) = e_1' \left(Z'_{-i,t'} W^h_{-i,t'} Z_{-i,t'} \right)^{-1} Z'_{-i,t'} W^h_{-i,t'} \Delta Y_{-i,t'}$, where the dependence of the weighting matrix W on the bandwidth h_n is now explicitly stated, and the bandwidth h_n minimizes the least squares criterion given by

$$CV(h_n) = \sum_{i=1}^{N} \sum_{t>t'} \left[\Delta Y_{it,t'} - \widehat{m}^{h}_{-t'}(Z_{it,t'}) \right]^2 . \tag{11}$$

For the linear model, as the scale effect exactly equals one, our average derivative estimators $\widehat{\delta}_{t'}$ defined in equation (6) could consistently estimate the parameters (not just up to scale) for $t' = 1, \ldots, (T-1)$. For the weighted average derivative estimators $\widehat{\delta}_W$ defined in equation (7), we adopt two different weighting functions:

$$\widehat{\delta}_{WStd} = \sum_{t'=1}^{T-1} \frac{\widehat{\delta}_{t'}/Std(\widehat{\delta}_{t'})}{\sum_{t'=1}^{T-1} 1/Std(\widehat{\delta}_{t'})}, \quad \widehat{\delta}_{WRMSE} = \sum_{t'=1}^{T-1} \frac{\widehat{\delta}_{t'}/RMSE(\widehat{\delta}_{t'})}{\sum_{t'=1}^{T-1} 1/RMSE(\widehat{\delta}_{t'})},$$

where $Std(\widehat{\delta}_{t'})$ and $RMSE(\widehat{\delta}_{t'})$ denote the standard deviation and root mean squared error of $\widehat{\delta}_{t'}$, respectively. The pairwise-difference average derivative estimator $\widehat{\delta}_{PW}$ defined in equation (9) is evaluated as well. The results are reported in Table 1. The third column indicates the true parameters, while the fourth to the last columns report the bias, RMSE, 2.5% quantile, median, and 97.5% quantile of the estimates, respectively. In the linear model, our estimators are compared with the stardard fixed effect estimator $\widehat{\delta}_{FElinear}$ using the within-group estimation procedure. While all estimators are practically unbiased, the RMSE of $\widehat{\delta}_{t'}$ is increasing with the time difference t'. This is not surprising as the number of observations after differencing decreases as t' grows and the first-difference estimator $\widehat{\delta}_1$ is thus most precise. Even smaller RMSEs are obtained by the weighted and pairwise-difference estimators. The RMSE of these average derivative estimators are roughly 30–50% larger than those of the within-group estimator.

For the binary-choice logit model, several alternative methods are reported for comparison. The first comparison is made with the conditional fixed-effect logit estimator $\widehat{\delta}_{FElogit}$. Furthermore, we consider the pairwise smoothed maximum score estimator $\widehat{\delta}^{SMS}_{PW}$ in Charlier et al. (1995). To make it comparable with the first-difference estimator $\widehat{\delta}_1$, we also include the first-difference smoothed maximum score estimator $\widehat{\delta}^{SMS}_1$. In all cases (even for the fixed-effect logit), estimates are normalized such that their norms equal to 1. The simulation results are summarized in Tables 2–4 for the homoscedastic, heteroscedastic,

Table 1: The bias, RMSE, and quartiles of all estimators in the CRE linear model.

	Parameters	True	Bias	RMSE	LQ	Median	UQ
$\widehat{\delta}_1$	β_1	1	0.0015	0.0589	0.8881	1.0035	1.1224
	β_2	-1	0.0033	0.0615	-1.1245	-0.9954	-0.8782
$\widehat{\delta}_2$	β_1	1	0.0008	0.0610	0.8844	1.0015	1.1289
	β_2	-1	0.0032	0.0633	-1.1217	-0.9963	-0.8708
$\widehat{\delta}_3$	β_1	1	-0.0009	0.0618	0.8800	0.9977	1.1181
	β_2	-1	0.0044	0.0678	-1.1278	-0.9933	-0.8612
$\widehat{\delta}_4$	β_1	1	0.0007	0.0739	0.8590	0.9999	1.1500
	β_2	-1	0.0070	0.0825	-1.1497	-0.9924	-0.8300
$\widehat{\delta}_5$	β_1	1	0.0028	0.0888	0.8249	1.0022	1.1742
	β_2	-1	0.0041	0.0934	-1.1789	-0.9974	-0.8113
$\widehat{\delta}_6$	β_1	1	-0.0033	0.1128	0.7790	0.9985	1.2172
	β_2	-1	-0.0012	0.1283	-1.2577	-0.9979	-0.7585
$\widehat{\delta}_7$	β_1	1	-0.0075	0.2186	0.5514	0.9942	1.4044
	β_2	-1	-0.0069	0.2331	-1.4854	-1.0065	-0.5519
$\widehat{\delta}_{WStd}$	β_1	1	0.0002	0.0505	0.9019	1.0000	1.0965
	β_2	-1	0.0032	0.0531	-1.1005	-0.9962	-0.8990
$\widehat{\delta}_{WRMSE}$	β_1	1	0.0002	0.0505	0.9019	1.0000	1.0965
	β_2	-1	0.0031	0.0531	-1.1005	-0.9963	-0.8990
$\widehat{\delta}_{PW}$	β_1	1	0.0022	0.0552	0.8879	1.0007	1.1075
	β_2	-1	0.0020	0.0558	-1.1122	-0.9976	-0.8879
$\widehat{\delta}_{FElinear}$	β_1	1	0.0006	0.0377	0.9293	0.9998	1.0731
	β_2	-1	0.0019	0.0379	-1.0717	-0.9984	-0.9260

Note: For the linear models, the estimators do not have to be scale normalized. LQ and UQ are 2.5% and 97.5% quantiles, respectively.

Table 2: The bias, RMSE, and quartiles of all estimators in the CRE logit model.

	Parameters	True	Bias	RMSE	LQ	Median	UQ
$\widehat{\delta}_1$	β_1	0.7071	-0.0073	0.0695	0.5548	0.7057	0.8259
	β_2	-0.7071	-0.0005	0.0685	-0.8320	-0.7085	-0.5639
$\widehat{\delta}_2$	β_1	0.7071	-0.0054	0.0725	0.5568	0.7069	0.8413
	β_2	-0.7071	0.0021	0.0733	-0.8306	-0.7073	-0.5406
$\widehat{\delta}_3$	β_1	0.7071	-0.0028	0.0778	0.5414	0.7079	0.8527
	β_2	-0.7071	0.0059	0.0798	-0.8408	-0.7063	-0.5223
$\widehat{\delta}_4$	β_1	0.7071	-0.0038	0.0884	0.5132	0.7062	0.8573
	β_2	-0.7071	0.0073	0.0889	-0.8582	-0.7080	-0.5148
$\widehat{\delta}_5$	β_1	0.7071	-0.0004	0.1060	0.4711	0.7125	0.8873
	β_2	-0.7071	0.0169	0.1099	-0.8821	-0.7017	-0.4612
$\widehat{\delta}_6$	β_1	0.7071	-0.0077	0.1370	0.4103	0.7088	0.9274
	β_2	-0.7071	0.0199	0.1426	-0.9120	-0.7054	-0.3742
$\widehat{\delta}_7$	β_1	0.7071	-0.0334	0.2633	0.0369	0.7316	0.9964
	β_2	-0.7071	0.0880	0.3203	-0.9923	-0.6753	0.2017
$\widehat{\delta}_{WStd}$	β_1	0.7071	-0.0002	0.0598	0.5836	0.7108	0.8208
	β_2	-0.7071	0.0049	0.0603	-0.8121	-0.7034	-0.5712
$\widehat{\delta}_{WRMSE}$	β_1	0.7071	-0.0003	0.0597	0.5834	0.7107	0.8206
	β_2	-0.7071	0.0048	0.0602	-0.8122	-0.7035	-0.5715
$\widehat{\delta}_{FElogit}$	β_1	0.7071	-0.0039	0.0457	0.6147	0.7042	0.7936
	β_2	-0.7071	-0.0009	0.0455	-0.7888	-0.7100	-0.6084
$\widehat{\delta}^{SMS}_{(1)}$	β_1	0.7071	-0.0481	0.1233	0.4688	0.6533	0.8839
	β_2	-0.7071	-0.0275	0.1182	-0.8833	-0.7571	-0.4677
$\widehat{\delta}^{SMS}_{PW}$	β_1	0.7071	-0.0290	0.1041	0.4851	0.6811	0.8796
	β_2	-0.7071	-0.0127	0.1106	-0.8744	-0.7322	-0.4756

Note: LQ and UQ are 2.5% and 97.5% quantiles, respectively. For comparison, all parameters, including true ones, are normalized such that $||\beta|| = 1$.

and random-effect logits, respectively. The ordering of various average-derivative estimators stays the same as in the case of the linear regression model. It achieves again RMSEs larger by 30–50% than the fixed-effects logit, which exhibits generally the smallest RMSE – even in the heteroscedastic model as its inconsistency influences the parameters in absolute values, but not after normalization. On the other hand, the first-difference and weighted average derivative estimates always outperform the smoothed maximum score estimation.

For the CRE Tobit model, we compare our method with the estimator $\widehat{\delta}_{Honore}$ obtained by trimmed least squares of Honoré (1992) and with the bias corrected Jack-knife estimator $\widehat{\delta}_{Jackknife}$ of Hahn and Newey (2004). Again, all methods deliver practically

Table 3: The bias, RMSE, and quartiles of all estimators in the CRE logit model with heteroscedasticity.

	Parameters	True	Bias	RMSE	LQ	Median	UQ
$\widehat{\delta}_1$	β_1	0.7071	-0.0095	0.0844	0.5173	0.7060	0.8402
	β_2	-0.7071	0.0003	0.0824	-0.8558	-0.7083	-0.5422
$\widehat{\delta}_2$	β_1	0.7071	-0.0087	0.0891	0.5194	0.7013	0.8572
	β_2	-0.7071	0.0026	0.0904	-0.8545	-0.7129	-0.5150
$\widehat{\delta}_3$	β_1	0.7071	-0.0071	0.0959	0.4919	0.7071	0.8776
	β_2	-0.7071	0.0061	0.0970	-0.8707	-0.7071	-0.4794
$\widehat{\delta}_4$	β_1	0.7071	-0.0070	0.1071	0.4734	0.7044	0.8854
	β_2	-0.7071	0.0094	0.1084	-0.8809	-0.7098	-0.4649
$\widehat{\delta}_5$	β_1	0.7071	-0.0027	0.1283	0.4242	0.7100	0.9370
	β_2	-0.7071	0.0223	0.1378	-0.9056	-0.7042	-0.3493
$\widehat{\delta}_6$	β_1	0.7071	-0.0132	0.1644	0.3315	0.7008	0.9653
	β_2	-0.7071	0.0291	0.1814	-0.9434	-0.7133	-0.2571
$\widehat{\delta}_7$	β_1	0.7071	-0.0727	0.3486	-0.2737	0.7176	0.9980
	β_2	-0.7071	0.1223	0.3930	-0.9929	-0.6754	0.4520
$\widehat{\delta}_{WStd}$	β_1	0.7071	-0.0022	0.0738	0.5509	0.7087	0.8368
	β_2	-0.7071	0.0055	0.0743	-0.8346	-0.7055	-0.5475
$\widehat{\delta}_{WRMSE}$	β_1	0.7071	-0.0023	0.0737	0.5512	0.7088	0.8367
	β_2	-0.7071	0.0054	0.0741	-0.8344	-0.7054	-0.5477
$\widehat{\delta}_{FElogit}$	β_1	0.7071	-0.0050	0.0556	0.5863	0.7042	0.8045
	β_2	-0.7071	-0.0007	0.0552	-0.8101	-0.7100	-0.5939
$\widehat{\delta}_1^{MS}$	β_1	0.7071	-0.0480	0.1327	0.4618	0.6477	0.9107
	β_2	-0.7071	-0.0236	0.1303	-0.8870	-0.7619	-0.4131
$\widehat{\delta}_{PW}^{MS}$	β_1	0.7071	-0.0293	0.1091	0.4847	0.6810	0.8865
	β_2	-0.7071	-0.0114	0.1161	-0.8747	-0.7323	-0.4628

Note: LQ and UQ are 2.5% and 97.5% quantiles, respectively. For comparison, all parameters, including true ones, are normalized such that $||\beta|| = 1$.

Table 4: The bias, RMSE, and quartiles of all estimators in the random effects logit model.

	Parameters	True	Bias	RMSE	LQ	Median	UQ
$\widehat{\delta}_1$	β_1	0.7071	-0.0033	0.0744	0.5424	0.7054	0.8398
	β_2	-0.7071	0.0047	0.0756	-0.8401	-0.7088	-0.5429
$\widehat{\delta}_2$	β_1	0.7071	0.0002	0.0721	0.5517	0.7091	0.8385
	β_2	-0.7071	0.0077	0.0739	-0.8340	-0.7051	-0.5450
$\widehat{\delta}_3$	β_1	0.7071	-0.0046	0.0829	0.5295	0.7059	0.8611
	β_2	-0.7071	0.0054	0.0848	-0.8483	-0.7083	-0.5085
$\widehat{\delta}_4$	β_1	0.7071	-0.0035	0.0966	0.4971	0.7080	0.8833
	β_2	-0.7071	0.0100	0.0992	-0.8677	-0.7062	-0.4688
$\widehat{\delta}_5$	β_1	0.7071	-0.0164	0.1166	0.4516	0.7001	0.8905
	β_2	-0.7071	0.0024	0.1140	-0.8922	-0.7140	-0.4549
$\widehat{\delta}_6$	β_1	0.7071	-0.0213	0.1548	0.3393	0.7038	0.9418
	β_2	-0.7071	0.0123	0.1537	-0.9407	-0.7104	-0.3362
$\widehat{\delta}_7$	β_1	0.7071	-0.0602	0.3095	-0.1038	0.7040	0.9962
	β_2	-0.7071	0.0865	0.3346	-0.9943	-0.6982	0.2467
$\widehat{\delta}_{WStd}$	β_1	0.7071	-0.0022	0.0640	0.5718	0.7056	0.8238
	β_2	-0.7071	0.0037	0.0652	-0.8204	-0.7086	-0.5669
$\widehat{\delta}_{WRMSE}$	β_1	0.7071	-0.0022	0.0639	0.5717	0.7054	0.8238
	β_2	-0.7071	0.0036	0.0651	-0.8205	-0.7088	-0.5669
$\widehat{\delta}_{PW}$	β_1	0.7071	-0.0053	0.0697	0.5629	0.7042	0.8313
	β_2	-0.7071	0.0017	0.0703	-0.8265	-0.7100	-0.5558
$\widehat{\delta}_{FElogit}$	β_1	0.7071	-0.0007	0.0484	0.6100	0.7066	0.7995
	β_2	-0.7071	0.0027	0.0491	-0.7924	-0.7076	-0.6007
$\widehat{\delta}_1^{MS}$	β_1	0.7071	-0.0371	0.1284	0.4637	0.6675	0.9149
	β_2	-0.7071	-0.0131	0.1325	-0.8860	-0.7446	-0.4037
$\widehat{\delta}_{PW}^{MS}$	β_1	0.7071	-0.0243	0.1051	0.4922	0.6817	0.8744
	β_2	-0.7071	-0.0079	0.1103	-0.8705	-0.7317	-0.4851

Note: LQ and UQ are 2.5% and 97.5% quantiles, respectively. For comparison, all parameters, including true ones, are normalized such that $||\beta|| = 1$.

Table 5: The bias, RMSE, and quartiles of all estimators in the CRE Tobit model.

	Parameters	True	Bias	RMSE	LQ	Median	UQ
$\widehat{\delta}_1$	β_1	0.7071	-0.0005	0.0396	0.6259	0.7085	0.7832
	β_2	-0.7071	0.0017	0.0398	-0.7799	-0.7057	-0.6127
$\widehat{\delta}_2$	β_1	0.7071	-0.0012	0.0422	0.6175	0.7073	0.7818
	β_2	-0.7071	0.0013	0.0420	-0.7865	-0.7069	-0.6235
$\widehat{\delta}_3$	β_1	0.7071	-0.0008	0.0446	0.6110	0.7090	0.7936
	β_2	-0.7071	0.0021	0.0450	-0.7916	-0.7053	-0.6084
$\widehat{\delta}_4$	β_1	0.7071	0.0003	0.0540	0.6006	0.7091	0.8068
	β_2	-0.7071	0.0045	0.0547	-0.7995	-0.7052	-0.5909
$\widehat{\delta}_5$	β_1	0.7071	0.0002	0.0629	0.5785	0.7094	0.8235
	β_2	-0.7071	0.0058	0.0638	-0.8157	-0.7048	-0.5673
$\widehat{\delta}_6$	β_1	0.7071	-0.0087	0.0879	0.5188	0.7049	0.8492
	β_2	-0.7071	0.0020	0.0868	-0.8549	-0.7093	-0.5280
$\widehat{\delta}_7$	β_1	0.7071	-0.0205	0.1574	0.3358	0.7038	0.9435
	β_2	-0.7071	0.0143	0.1565	-0.9419	-0.7104	-0.3313
$\widehat{\delta}_{WStd}$	β_1	0.7071	-0.0004	0.0363	0.6313	0.7081	0.7742
	β_2	-0.7071	0.0015	0.0363	-0.7756	-0.7061	-0.6330
$\widehat{\delta}_{WRMSE}$	β_1	0.7071	-0.0004	0.0363	0.6314	0.7081	0.7741
	β_2	-0.7071	0.0015	0.0364	-0.7755	-0.7061	-0.6330
$\widehat{\delta}_{Honore}$	β_1	0.7071	-0.0006	0.0260	0.6535	0.7066	0.7557
	β_2	-0.7071	0.0003	0.0260	-0.7570	-0.7076	-0.6549
$\widehat{\delta}_{Jackknife}$	β_1	0.7071	0.0001	0.0230	0.6616	0.7074	0.7512
	β_2	-0.7071	0.0008	0.0230	-0.7499	-0.7068	-0.6600

Note: LQ and UQ are 2.5% and 97.5% quantiles, respectively. For comparison, all parameters, including true ones, are normalized such that $||\beta|| = 1$.

unbiased estimates and the average-derivative estimator exhibits again RMSEs that are approximately 40% larger than those of the methods specialized to the censored regression models.

Altogether, the average derivative estimators (both the first-differenced and weighted forms) deliver a robust performance across a range of nonlinear panel data models. Although they do not reach the precision of the estimation methods specialized to each particular model, they offer much wider applicability than the methods specifically designed for one kind of model and open up possibility to estimate many new panel data models with correlated random effects.

6 Conclusion

Both regression coefficients and marginal effects in nonseparable single-index models with correlated random effects are shown to be identified. The suggested estimation procedure relies on the local polynomial regression and the average derivative estimation. The estimation of the slope coefficients requires only two time periods and is not only consistent and asymptotically, but also exhibits reasonably good finite sample performance in a variety of panel data models. The procedure is currently limited to static panel data models and an extension to dynamic panel data is a topic of future research.

A Proof of Theorem 1

This appendix provides the proof of the main identification result.

Proof of Theorem 1. Under fairly general conditions, expectation $E[Y_{it}|X_{it}, X_{i(t-t')}]$ and its derivatives are nonparametrically identified from the observed data and can be written at $(x_t, x_{t-t'})$ as

$$
\begin{aligned}
&E[Y_{it}|X_{it} = x_t, X_{i(t-t')} = x_{t-t'}] = E[\phi(X_{it}'\beta, \alpha, u_t)|X_{it} = x_t, X_{i(t-t')} = x_{t-t'}] \\
&= \int \phi(x_t'\beta, \alpha, u_t) \times F_{U_t,\alpha|X_t,X_{t-t'}}(du_t, d\alpha|X_{it} = x_t, X_{i(t-t')} = x_{t-t'}).
\end{aligned}
$$

To simplify the notation, we write $F(u_t, \alpha|x_t, x_{t-t'}) \equiv F_{U_t,\alpha|X_t,X_{t-t'}}(u_t, \alpha|X_{it} = x_t, X_{i(t-t')} = x_{t-t'})$. Further, $f(\alpha|x_t, x_{t-t'}) \equiv f_{\alpha|X_t,X_{t-t'}}(\alpha|X_{it} = x_t, X_{i(t-t')} = x_{t-t'})$. Applying successive conditioning leads to

$$
\begin{aligned}
&E[Y_{it}|X_{it} = x_t, X_{i(t-t')} = x_{t-t'}] \\
&= \int \left[\int \phi(x_t'\beta, \alpha, u_t) F_{U|\alpha,X_t,X_{t-t'}}(du_t|\alpha, x_t, x_{t-t'}) \right] f(\alpha|x_t, x_{t-t'}) d\alpha \\
&= \int \left[\int \phi(x_t'\beta, \alpha, u_t) F_U(du_t) \right] f(\alpha|x_t, x_{t-t'}) d\alpha,
\end{aligned}
$$

where F_U denotes the distribution function of U_{it}, which is independent of α_i, X_{it}, and $X_{i(t-t')}$ by Assumption 2, and $f(\alpha|x_t, x_{t-t'}) \equiv f_{\alpha|X_t,X_{t-t'}}(\alpha|X_{it} = x_t, X_{i(t-t')} = x_{t-t'})$ denotes the conditional density of α.

As Assumptions 4, 5, and 6 ensure that the derivatives of this expectation exist and the valid interchange the order of integration and derivative, it follows that

$$
\begin{aligned}
\frac{\partial}{\partial x_t} E[Y_{it}|X_{it} = x_t, X_{i(t-t')} = x_{t-t'}] = &\int \int \left[\frac{\partial}{\partial x_t} \phi(x_t'\beta, \alpha, u_t) \right] F_U(du_t) f(\alpha|x_t, x_{t-t'}) d\alpha \\
&+ \int \left[\int \phi(x_t'\beta, \alpha, u_t) F_U(du_t) \right] \frac{\partial}{\partial x_t} f(\alpha|x_t, x_{t-t'}) d\alpha.
\end{aligned}
\tag{A.1}
$$

As $\frac{\partial}{\partial x_t}\phi(x_t'\beta, \alpha, u_t) = \phi_{xb}^{'}(x_t'\beta, \alpha, u_t)\beta$, the first part of the right handside of the above equation (A.1) can be rewritten as

$$\beta \int \int \phi_{xb}^{'}(x_t'\beta, \alpha, u_t) F_U(du_t) f(\alpha|x_t, x_{t-t'}) d\alpha = \beta E\left[\phi_{xb}^{'}(X_{it}'\beta, \alpha_i, U_{it}) | X_{it} = x_t, X_{i(t-t')} = x_{t-t'}\right]. \tag{A.2}$$

The second part of the above equation (A.1) can be expressed, by Assumption 3 implying $f(\alpha|x_t, x_{t-t'}) = f(\alpha|x_t + x_{t-t'})$, as

$$\int \left[\int \phi(x_t'\beta, \alpha, u_t) F_U(du_t) \right] f'(\alpha|x_t + x_{t-t'}) d\alpha. \tag{A.3}$$

The marginal effects of the covariates on the above conditional expectation (A.1) thus consist of two parts. The first part (A.2) represents the direct effect of a change in X_{it} averaged over the individual unobserved heterogeneity, whereas the second part (A.3) reflects the effect of a change in α_i on Y_{it} that caused by the change of X_{it}. However, when considering the marginal effects of the value X_{it} on past $Y_{i(t-t')}$, the first part of the effect disappears as $Y_{i(t-t')}$ does not depend on X_{it}. Therefore, we can write

$$\frac{\partial}{\partial x_t} E[Y_{i(t-t')} | X_{it} = x_t, X_{i(t-t')} = x_{t-t'}] = \int \left[\int \phi(x_{t-t'}'\beta, \alpha, u_{t-t'}) F_U(du_{t-t'}) \right] \frac{\partial}{\partial x_t} f(\alpha|x_t, x_{t-t'}) d\alpha \tag{A.4}$$

$$= \int \left[\int \phi(x_{t-t'}'\beta, \alpha, u_{t-t'}) F_U(du_{t-t'}) \right] f'(\alpha|x_t + x_{t-t'}) d\alpha. \tag{A.5}$$

Since X_{it} and $X_{i(t-t')}$ are identically distributed, integrating the conditional expectations (A.3) and (A.5) leads to the same quantity:

$$E_{X_t, X_{t-t'}} \left\{ \int \left[\int \phi(x_{t-t'}'\beta, \alpha, u_{t-t'}) F_U(du_{t-t'}) \right] f'(\alpha|x_t + x_{t-t'}) d\alpha \right\}$$
$$= E_{X_t, X_{t-t'}} \left\{ \int \left[\int \phi(x_t'\beta, \alpha, u_t) F_U(du_t) \right] f'(\alpha|x_t + x_{t-t'}) d\alpha \right\}.$$

This result implies that (A.1) can be rewritten using (A.2) as

$$
\begin{aligned}
&E\left\{\frac{\partial}{\partial x_t}E[Y_{it}|X_{it}=x_t, X_{i(t-t')}=x_{t-t'}]\right\}\\
=&\beta E\left[\phi'_{xb}(X_{it}'\beta,\alpha_i,U_{it})|X_{it}=x_t, X_{i(t-t')}=x_{t-t'}\right]-E\left\{\frac{\partial}{\partial x_t}E[Y_{i(t-t')}|X_{it}=x_t, X_{i(t-t')}=x_{t-t'}]\right\}.
\end{aligned}
$$

Therefore, similar to Härdle and Stoker (1989), β could be identified up to scale and the average estimator can be based on

$$
\begin{aligned}
\delta_{t'}=\gamma_{t'}\beta=\beta E\left\{E\left[\phi'_{xb}(X_{it}'\beta,\alpha_i,U_{it})|X_{it},X_{i(t-t')}\right]\right\}\\
=E\left\{\frac{\partial}{\partial X_{it}}E[Y_{it}-Y_{i(t-t')}|X_{it},X_{i(t-t')}]\right\},
\end{aligned}
$$

where $\gamma_{t'}=E\left(E\left[\phi'_{xb}(X_{it}'\beta,\alpha_i,U_{it})|X_{it},X_{i(t-t')}\right]\right)$ is a scalar (assumed to be nonzero).

Finally, when $\phi(X_t'\beta,\alpha,U_t)=X_t'\beta+\psi(\alpha,U_t)$, we have $\phi'_{xb}(X_t'\beta,\alpha,U_t)=1$ and $\gamma_{t'}=1$ for all t'. In this case, β is point identified for each t'. $\square$

B Proofs of Theorems 3 and 4

For the proofs of Theorems 3 and 4, we need to introduce notation, which is closely related to Masry (1996) and Li et al. (2003). Moreover, note that Assumptions 1–7 cover all assumptions used in Li et al. (2003) and Masry (1996) so that their results regarding the local polynomial estimator can be applied in the current context. After introducing the notation and some auxiliary lemmas, the proofs of the main theorems follow.

First, assuming that $m(\cdot)$ has $p+1$ derivatives at point z_0, we can approximate $m(z)$ locally by a multivariate polynomial of order p:

$$
m(z)\approx\sum_{0\le|k|\le p}\frac{1}{k!}D^k m(v)|_{v=z_0}(z-z_0)^k,
$$

where $k=(k_1,\ldots,k_{2d})$, $k!=k_1!\times\cdots\times k_{2d}!$, $|k|=\sum_{i=1}^{2d}k_i$, $z^k=z_1^{k_1}\times\cdots\times z_{2d}^{k_{2d}}$ and

$$
\sum_{0\le|k|\le p}=\sum_{j=0}^{p}\sum_{k_1=0}^{j}\cdots\sum_{k_{2d}=0;k_1+\cdots+k_{2d}=j},\quad\text{and}\quad (D^k m)(z)=\frac{\partial^k m(z)}{\partial z_1^{k_1}\ldots\partial z_{2d}^{k_{2d}}}.
$$

Further define $(V_{it,t'} = \Delta Y_{it,t'} - m(Z_{it,t'}))$

$$\begin{aligned}\bar{\tau}_{t',j}(z) &= \frac{1}{n(T-t')}\sum_{i=1}^{n}\sum_{t>t'}^{T}\left(\Delta Y_{it,t'} - m(Z_{it,t'})\right)\left(\frac{Z_{it,t'}-z}{h_n}\right)^{j} K_h\left(Z_{it,t'}-z\right) \\ &= \frac{1}{n(T-t')}\sum_{i=1}^{n}\sum_{t>t'}^{T} V_{it,t'}\left(\frac{Z_{it,t'}-z}{h_n}\right)^{j} K_h\left(Z_{it,t'}-z\right),\end{aligned}$$

and

$$\bar{s}_{t',j}(z) = \frac{1}{n(T-t')}\sum_{i=1}^{n}\sum_{t>t'}^{T}\left(\frac{Z_{it,t'}-z}{h_n}\right)^{j} K_h\left(Z_{it,t'}-z\right), \quad K_h(u) = \frac{1}{h_n^{2d}}K\left(\frac{u}{h_n}\right).$$

Following Masry (1996) and Li et al. (2003), we write $\bar{\tau}_{t',j}$ in a matrix form by using a lexicographical order in the following manner. Let

$$N_i = \begin{pmatrix} i+2d-1 \\ 2d-1 \end{pmatrix}$$

be the number of distinct $2d$-tuples with $|j| \equiv j_1 + \cdots + j_{2d} = i$ (N_i is the number of distinct derivatives of total order i). These N_i $2d$-tuples will be arranged as a sequence in a lexicographical order with the highest priority to the last position so that $(0,\ldots,0,i)$ is the first element in the sequence and $(i,0,\ldots,0)$ is the last element, and let $g_{|j|}^{-1}$ denote this one-to-one map. Arrange the $N_{|j|}$ values of the $\bar{\tau}_{t',j}$ in a column vector $\bar{\tau}_{t',|j|}$ according to this order. Then $(\tau_{t',|j|})_k = \bar{\tau}_{t',g_{|j|}}(k)$. Define $\tau_{t'} = (\tau_{t',0}', \tau_{t',1}', \ldots, \tau_{t',p}')'$, where $\tau_{t',i}$ is a $N_i \times 1$ vector with elements of $\bar{\tau}_{t',j}(z)$ arranged in the above lexicographical order. Note that $\tau_{t'}$ is of dimension $N \times 1$ with $N = \sum_{i=0}^{p} N_i$. Similarly, column vector $m_{p+1}(z)$ denotes the N_{p+1} elements of derivatives $(1/j!(D^j m))(z)$ for $|j| = p+1$ using the same lexicographical order

Next, the possible values of $\bar{s}_{t',j+k}$ are also arranged in a matrix $S_{t',|j|,|k|}$ in a lexicographical order with the (l,m)th element $[S_{t',|j|,|k|}]_{lm} = \bar{s}_{t',g_{|j|}(l)+g_{|k|}(m)}$. Now define

$$S_{t'} = \begin{pmatrix} S_{t',0,0} & S_{t',0,1} & \dots & S_{t',0,p} \\ S_{t',1,0} & S_{t',1,1} & \dots & S_{t',1,p} \\ \vdots & \vdots & \ddots & \vdots \\ S_{t',p,0} & S_{t',p,1} & \dots & S_{t',p,p} \end{pmatrix} \quad \text{and} \quad B_{t'}(z) = \begin{pmatrix} S_{t',0,p+1} \\ S_{t',1,p+1} \\ \vdots \\ S_{t',p,p+1} \end{pmatrix}.$$

Similar matrices are defined also for kernel moments $\mu_j = \int_{\mathbb{R}^{2d}} u^j K(u)du$ and $v_{s,j} = \int_{\mathbb{R}^{2d}} u_s u^j K(u)du$, where u_s is the sth component of vector u. Thus, let $M_{i,j}$ and $Q_{s,i,j}$ be $N_i \times N_j$ dimensional matrices whose (l,m)-th elements are given by $\mu_{g_i(l)+g_j(m)}$ and $v_{s,g_i(l)+g_j(m)}$, respectively, $s = 1,\dots,2d$, and let

$$M = \begin{pmatrix} M_{0,0} & M_{0,1} & \dots & M_{0,p} \\ M_{1,0} & M_{1,1} & \dots & M_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ M_{p,0} & M_{p,1} & \dots & M_{p,p} \end{pmatrix}, B = \begin{pmatrix} M_{0,p+1} \\ M_{1,p+1} \\ \vdots \\ M_{p,p+1} \end{pmatrix}, \text{and}\, Q_s = \begin{pmatrix} Q_{s,0,0} & Q_{s,0,1} & \dots & Q_{s,0,p} \\ Q_{s,1,0} & Q_{s,1,1} & \dots & Q_{s,1,p} \\ \vdots & \vdots & \ddots & \vdots \\ Q_{s,p,0} & Q_{s,p,1} & \dots & Q_{s,p,p} \end{pmatrix}.$$

Finally, we define $M^f(z) = Mf(z)$ and $Q^f(z) = \sum_{s=1}^{2d} f_s^{'}(z)Q_s$, where $f_s^{'}(z)$ is the s-th component of the first derivative $f^{'}(z)$ of the density function $f(z)$ of $Z_{it,t'}$, $s = 1,\dots,2d$.

Using this notation, Masry (1996), equation (2.13), and Li et al. (2003), equation (A.9), showed that

$$\widehat{\beta}_{t'}(z) - \beta(z) = S_{t'}^{-1}(z)\tau_{t'}(z) + h_n^{p+1}S_{t'}^{-1}B_{t'}(z)m_{p+1}(z) + o_p(h_n^{p+1}), \quad 0 \le |k| \le p, \quad \text{(A.6)}$$

where $\widehat{\beta}_{t'} = (\widehat{\beta}'_{t',0}, \widehat{\beta}'_{t',1}, \dots, \widehat{\beta}'_{t',p})'$, $\widehat{\beta}_{t',k} = h_n^{|k|}\widehat{b}_{k,t'}$, and $\widehat{b}_{k,t'}$ are the estimates of parameters $b_{k,t'}$ in objective function (4). Note that the subscript n denoting the cross-sectional dimension is kept implicit as there are many other subscripts needed already. The limits are taken and all asymptotic statements are stated for $n \to +\infty$ as T is fixed.

Further recall that our local derivative estimator is defined as the first d elements of $\widehat{\delta}_{t'}(z) = h_n^{-1}\widehat{\beta}_{t',1}(z) = h_n^{-1}L\widehat{\beta}_{t'}(z) = L\widehat{b}_{t'}(z)$ as in equation (5) and the average derivative estimator is defined as in equation (6) by

$$\widehat{\delta}_{t'} = \frac{1}{n(T-t')} \sum_{i=1}^{n} \sum_{t=t'+1}^{T} \widehat{\delta}_{t'}(Z_{it,t'}) = \frac{1}{n(T-t')h} \sum_{i=1}^{n} \sum_{t=t'+1}^{T} L\widehat{\beta}_{t'}(Z_{it,t'}).$$

To simplify notation, let $\mathcal{I}_{t'}$ denote the index set $\{it, t'\}_{i=1,t=t'+1}^{n\;\;;T}$. Sorting the sequence $\{Z_{it,t'}\}$ by the cross-sectional and the time indices, it is possible to express double sums with respect to i and t as $\sum_{l\in\mathcal{I}_{t'}} Z_l = \sum_{i=1}^{n}\sum_{t>t'}^{T} Z_{it,t'}$, or with a slight abuse of notation, by $\sum_{l=1}^{n(T-t')} Z_l$.

To derive the asymptotic distribution of $\widehat{\delta}_{t'}$ in Theorem 3, we will consider the following sample average of $\widehat{\beta}_{t'} - \beta$ as in Li et al. (2003), substituting from (A.6),

$$\begin{aligned} &\frac{1}{n(T-t')} \sum_{l\in\mathcal{I}_{t'}} [\widehat{\beta}_{t'}(Z_l) - \beta(Z_l)] \\ &= \frac{1}{n(T-t')} \sum_{l\in\mathcal{I}_{t'}} S_{t'}^{-1}(Z_l)\tau_{t'}(Z_l) + \frac{h_n^{p+1}}{n(T-t')} \sum_{l\in\mathcal{I}_{t'}} S_{t'}^{-1} B_{t'}(Z_l) m_{p+1}(Z_l) + o_p(h_n^{p+1}) \qquad \text{(A.7)} \\ &= A_{t'}^1 + h_n^{p+1} A_{t'}^2 + o(h_n^{p+1}), \end{aligned}$$

where $A_{t'}^1 = \frac{1}{n(T-t')}\sum_{l\in\mathcal{I}_{t'}} S_{t'}^{-1}(Z_l)\tau_{t'}(Z_l)$ and $A_{t'}^2 = \frac{1}{n(T-t')}\sum_{l\in\mathcal{I}_{t'}} S_{t'}^{-1}(Z_l)B_{t'}(Z_l)m_{p+1}(Z_l)$. Additionally, since Lemma A.1 of Li et al. (2003) holds for the strongly mixing processes and implies $S_{t'}^{-1}(z) = (M^f(z))^{-1} - h_n G(z) + o(h_n)$ a.s. uniformly in $z \in \mathcal{D}$, where $G^f(z) = (M^f(z))^{-1}Q^f(z)(M^f(z))^{-1}$, elements of $A_{t'}^1$ can be further decomposed to[1]

$$\begin{aligned} A_{t'}^1 &= \frac{1}{n(T-t')} \sum_{l\in\mathcal{I}_{t'}} \Big[e_r'(M^f(Z_l))^{-1}\tau_{t'}(Z_l) - h_n e_r' G^f(Z_l)\tau_{t'}(Z_l) \Big] + (s.o.) \\ &\equiv J_{t',r}^1 - hJ_{t',r}^2 + (s.o.), \end{aligned} \qquad \text{(A.8)}$$

where $J_{t',r}^1 = \frac{1}{n(T-t')}\sum_{l\in\mathcal{I}_{t'}} e_r'(M^f(Z_l))^{-1}\tau_{t'}(Z_l)$ and $J_{t',r}^2 = \frac{1}{n(T-t')}\sum_{l\in\mathcal{I}_{t'}} e_r' G^f(Z_l)\tau_{t'}(Z_l)$.

The following lemmas state the properties of terms $A_{t'}^2$, $J_{t',r}^1 = e_r' J_{t'}^1$, and $J_{t',r}^2 = e_r' J_{t'}^2$; again, all asymptotic statements are for $n \to +\infty$ and T being fixed.

Lemma 1. *$A_{t'}^2 = A_{t'} + O(h_n)$ a.s., where $A_{t'} = M^{-1}BE[m_{p+1}(Z_{it,t'})]$.*

Proof. Lemma A.1 by Li et al. (2003) again implies that $\sup_{z\in\mathcal{D}}|(s_{t'}(z))^{-1} - (M^f(z))^{-1}| =$

[1] We write $\mathcal{A}_n = \mathcal{B}_n + (s.o.)$ to denote the fact that $\mathcal{B}_n$ is the leading term of $\mathcal{A}_n$, $(s.o.)$ stands for terms that have smaller order than $\mathcal{B}_n$.

$O(h_n)$ a.s. and Masry (1996) have shown that $\sup_{z\in\mathcal{D}}|B_{t'}(z) - Bf(z)| = O(h_n)$ a.s. Thus, $A^2_{t'} = \frac{1}{n(T-t')} M^{-1} B \sum_{l\in\mathcal{I}_{t'}} m_{p+1}(Z_l) + O(h_n)$ a.s. As Assumption 7.4 guarantees that $m_{p+1}(.)$ is bounded and uniformly continuous, $m_{p+1}(Z_l)$ forms T-dependent sequence and thus a stationary and strong mixing process. Therefore, by the Corollary of Blum et al. (1963), $\frac{1}{n(T-t')}\sum_{i=1}^{n}\sum_{t>t'} m_{p+1}(Z_{it,t'})$ converges to $E[m_{p+1}(Z_{it,t'})]$ almost surely. Thus, $A^2_{t'} = M^{-1}BE[m_{p+1}(Z_{it,t'})] + O(h_n) = A_{t'} + O(h_n)$ a.s. □

Lemma 2. *$J^1_{t',r} = O_p\left((nh_n^d)^{-1}\right)$ for $r = 2, \ldots, 2d+1$.*

Proof. This is verified by Li et al. (2003, Lemma A.3). □

Lemma 3. *$J^2_{t',r} \to N(0, \Phi_{t',r})$ in distribution for $r = 2, \ldots, 2d+1$, where*

$$\Phi_{t',r} = \frac{1}{(T-t')^2}\sum_{t=t'+1}^{T}\sum_{s=t'+1}^{T} E\left[\sigma_{ts,t'}(Z_{it,t'}, Z_{is,t'})(G(Z_{it,t'}))_{r,1}(G(Z_{is,t'}))_{r,1}\right]$$

and matrix $G(z) = G^f(z)M^f(z) = [M^f(z)]^{-1}Q^f(z)$.

Proof. The proof closely follows Li et al. (2003, Lemma A.4). Denote $V_{it,t'} = \Delta Y_{it,t'} - m(Z_{it,t'})$; then

$$\begin{aligned}
J^2_{t',r} =& \frac{1}{n(T-t')}\sum_{l\in\mathcal{I}_{t'}} e_r' G^f(Z_l)\tau_{t'}(Z_l)\\
=& \frac{1}{n(T-t')}\sum_{l\in\mathcal{I}_{t'}}\sum_{0\le|j|\le p}(G^f(Z_l))_{r,j}\tau_{t',j}(Z_l)\\
=& \left[\frac{1}{n(T-t')}\right]^2\sum_{l\in\mathcal{I}_{t'}}\sum_{i\in\mathcal{I}_{t'}}\sum_{0\le|j|\le p} V_i(G^f(Z_l))_{r,j}\left(\frac{Z_i-Z_l}{h_n}\right)^j K_h(Z_i-Z_l)\\
=& \left[\frac{1}{n(T-t')}\right]^2\sum_{l\in\mathcal{I}_{t'}}\sum_{i\in\mathcal{I}_{t'},i\ne l}\sum_{0\le|j|\le p} V_i(G^f(Z_l))_{r,j}\left(\frac{Z_i-Z_l}{h_n}\right)^j K_h(Z_i-Z_l)\\
&+\left[\frac{1}{n(T-t')}\right]^2\sum_{i\in\mathcal{I}_{t'}} V_i(G^f(Z_i))_{r,0}K_h(0)\\
=& 2\left[\frac{1}{n(T-t')}\right]^2\sum_{l=1}^{n(T-t')}\sum_{i>l}^{n(T-t')} H_{t',r}(Z_i, Z_l) + O_p((n^{3/2}h_n^{2d})^{-1}),
\end{aligned}$$

where $(n^{3/2}h_n^{2d})^{-1} \to 0$ by Assumption (7).1 and the symmetrized elements

$$H_{t',r}(V_i, Z_i; V_l, Z_l) = \sum_{0\le|j|\le p}\frac{1}{2}\left[V_i(G^f(Z_l))_{r,j}\left(\frac{Z_i-Z_l}{h_n}\right)^j + V_l(G^f(Z_i))_{r,j}\left(\frac{Z_l-Z_i}{h_n}\right)^j\right]K_h(Z_i-Z_l).$$

Let for $i, l \in \mathcal{I}_{t'}$

$$\mathcal{H}_{t',r}(V_i, Z_i) = E[H_{t',r}(Z_i, Z_l)|V_i, Z_i] = \frac{1}{2}\sum_{0\le|j|\le p} V_i E\left[(G^f(Z_l))_{r,j}\left(\frac{Z_i - Z_l}{h_n}\right)^j K_h(Z_i - Z_l)|Z_i\right].$$

As $E\left[(G^f(Z_l))_{r,j}\left(\frac{Z_i-Z_l}{h_n}\right)^j K_h(Z_i - Z_l)|Z_i\right] = \int (G^f(z_l))_{r,j} f(z_l)\left(\frac{Z_i-z_l}{h_n}\right)^j K_h(Z_i - z_l)dz_l = \int (G^f(Z_i+h_n u))_{r,j} u^j f(Z_i+h_n u)K(u)du = (G^f(Z_i))_{r,j}\int u^j K(u)du f(Z_i)+O(h_n) = (G^f(Z_i))_{r,j}\mu_j f(Z_i)+O(h_n)$, it follows that $\mathcal{H}_{t',r}(V_i, Z_i) = \frac{1}{2}V_i\sum_{0\le|j|\le p}(G^f(Z_i))_{r,j}\mu_j f(Z_i) = \frac{1}{2}V_i(G^f(Z_i)M^f(Z_i))_{r,1} = \frac{1}{2}V_i(G(Z_i))_{r,1}$, $i \in \mathcal{I}_{t'}$. Therefore, by the U-statistics H-decomposition we have

$$J^2_{t',r} = \frac{2}{n(T-t')}\sum_{i\in\mathcal{I}_{t'}}\mathcal{H}_{t',r}(V_i, Z_i)+(s.o.) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{T-t'}\sum_{t>t'}^{T} V_{it,t'}(G(Z_{it,t'}))_{r,1}\right]+(s.o.). \tag{A.9}$$

Therefore, by the Lindenberg central limit theorem, $\sqrt{n}J^2_{t',r} \to N(0, \Phi_{t',r})$ for $r = 2, \dots, 2d+1$, where

$$\begin{aligned}\Phi_{t',r} &= Var\left[\frac{1}{T-t'}\sum_{t>t'}^{T} V_{it,t'}(G(Z_{it,t'}))_{r,1}\right] \\ &= \frac{1}{(T-t')^2}\sum_{t=t'+1}^{T}\sum_{s=t'+1}^{T} E\left[\sigma_{ts,t'}(Z_{it,t'}, Z_{is,t'})(G(Z_{it,t'}))_{r,1}(G(Z_{is,t'}))_{r,1}\right].\end{aligned}$$

□

Lemma 4. *Define $d \times 1$ vectors $J^2_{t',[d]} = LJ^2_{t'} = (J^2_{t',2}, \dots, J^2_{t',d+1})'$ and $(G(z))_{[d],1} = LG(z)e_1 = ((G(z))_{2,1}, \dots, (G(z))_{d+1,1})'$, where $(G(z))_{r,1}$ is the $(r,1)$-th element of $G(z) = [M^f(z)]^{-1}Q^f(z)$. Then $\sqrt{n}J^2_{t',[d]} \to N(0, \Phi_{t'})$, where*

$$\Phi_{t'} = \frac{1}{(T-t')^2}\sum_{t=t'+1}^{T}\sum_{s=t'+1}^{T} E\left[\sigma_{ts,t'}(Z_{it,t'}, Z_{is,t'})(G(Z_{it,t'}))_{[d],1}[(G(Z_{is,t'}))_{[d],1}]'\right].$$

Proof. By the equation (A.9) in the proof of Lemma 3, we know that

$$
\begin{aligned}
& Cov(\sqrt{n}J^2_{t',r}, \sqrt{n}J^2_{t',m}) \\
=& \frac{1}{n}\left[\frac{1}{T-t'}\right]^2 E\left[\left(\sum_{i=1}^{n}\sum_{t>t'}^{T} V_{it,t'}(G(Z_{it,t'}))_{r,1}\right)\left(\sum_{j=1}^{n}\sum_{s>t'}^{T} V_{js,t'}(G(Z_{js,t'}))_{m,1}\right)\right] + (s.o.) \\
=& \frac{1}{n}\left[\frac{1}{T-t'}\right]^2 \sum_{i=1}^{n} E\left[\left(\sum_{t>t'}^{T} V_{it,t'}(G(Z_{it,t'}))_{r,1}\right)\left(\sum_{s>t'}^{T} V_{is,t'}(G(Z_{is,t'}))_{m,1}\right)\right] + (s.o.) \\
=& \frac{1}{(T-t')^2}\sum_{t=t'+1}^{T}\sum_{s=t'+1}^{T} E\left[\sigma_{ts,t'}(Z_{it,t'}, Z_{is,t'})(G(Z_{it,t'}))_{r,1}(G(Z_{is,t'}))_{m,1}\right] + (s.o.) \\
=& (\Phi_{t'})_{r-1,m-1} + o(1).
\end{aligned}
$$

where $r, m = 2, \ldots, d+1$. Hence, $Var(\sqrt{n}J^2_{1,t'}) = \Phi_{t'} + o(1)$. Analogously to Li et al. (2003, Lemma A.5), one can easily show that the result of Lemma 3 and the Crammer-Wold device imply $\sqrt{n}J^2_{1,t'} \to N(0, \Phi_{t'})$. □

Proof of Theorem 3: Let $\tilde{\delta}_{t'} = \frac{1}{n(T-t')}\sum_{i=1}^{n}\sum_{t>t'} m_1'(Z_{it,t'})$ and define $d\times 1$ vectors $A_{t',[d]} = LA_{t'}$, $A^k_{t',[d]} = LA^k_{t'}$, and $J^k_{t',[d]} = LJ^k_{t'}$ (recalling that $L = (e_2, \ldots, e_{d+1})'$), where $k = 1, 2$ and $A_{t',j} = e_j' A_{t'}$, $A^k_{t',j}$, and $J^k_{t',j}$ are defined in (1), (A.7)–(A.8), $j = 2, \ldots, d+1$. By the results of Lemmas 2 and 4, and from equations (6), (A.7), and (A.8), we have

$$
\begin{aligned}
& \sqrt{n}\left(\widehat{\delta}_{t'} - \tilde{\delta}_{t'} - h_n^p A_{t',[d]}\right) \\
=& \frac{\sqrt{n}}{h}\left(\frac{1}{n(T-t')}\sum_{l\in\mathcal{I}_{t'}}[L\widehat{\beta}_{t'}(Z_l) - L\beta(Z_l) - h_n^{p+1}A_{1,t'}]\right) \\
=& \frac{\sqrt{n}}{h}\left(A^1_{t',[d]} + h^{p+1}A^2_{t',[d]} - h_n^{p+1}A_{t',[d]} + o_p(h_n^{p+1})\right) \qquad \text{(A.10)} \\
=& \frac{\sqrt{n}}{h}\left[J^1_{t',[d]} - hJ^2_{t',[d]} + h_n^{p+1}(A^2_{t',[d]} - A_{t',[d]}) + o_p(h_n^{p+1})\right] \\
=& O_p((nh_n^{d+2})^{-1/2}) - \sqrt{n}J^2_{t',[d]} + O_p(n^{1/2}h_n^{p+1}) \\
=& -\sqrt{n}J^2_{t',[d]} + o_p(1) \to N(0, \Phi_{t'}),
\end{aligned}
$$

where the last equality follows from Assumption 7.1. Under Assumption 7, and by the

Lindenberg central limit theorem, it also follows that

$$\begin{aligned}
&\sqrt{n}\left[\tilde{\delta}_{t'} - E\left(m_1'(Z_{it,t'})\right)\right] \\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\frac{1}{T-t'}\sum_{t>t'}^{T} m_1'(Z_{it,t'}) - E\left(m_1'(Z_{it,t'})\right)\right] \to N(0,\Omega_{t'}),
\end{aligned}$$

where

$$\Omega_{t'} = \frac{1}{(T-t')^2}\sum_{t=t'+1}^{T}\sum_{s=t'+1}^{T} Cov\left[m_1'(Z_{it,t'}), m_1'(Z_{is,t'})\right].$$

Similarly to Li et al. (2003), it can be shown that

$$\begin{aligned}
&Cov\left(\sqrt{n}(\widehat{\delta}_{t'} - \tilde{\delta}_{t'} - h_n^p A_{t',[d]}), \sqrt{n}\left[\tilde{\delta}_{t'} - E\left(m_1'(Z_{it,t'})\right)\right]\right) \\
=&\frac{1}{n}Cov\left(\frac{1}{T-t'}\sum_{l\in\mathcal{I}_{t'}}\left[\widehat{m}_1'(Z_l) - m_1'(Z_l) - h_n^p A_{t',[d]}\right], \frac{1}{T-t'}\sum_{l\in\mathcal{I}_{t'}}\left[m_1'(Z_l) - E\left(m_1'(Z_{it,t'})\right)\right]\right) \\
=&\frac{1}{n}E\left(\frac{1}{T-t'}\sum_{l\in\mathcal{I}_{t'}}\left[\widehat{m}_1'(Z_l) - m_1'(Z_l) - h_n^p A_{t',[d]}\right]\cdot\frac{1}{T-t'}\sum_{l\in\mathcal{I}_{t'}}\left[m_1'(Z_l) - E\left(m_1'(Z_{it,t'})\right)\right]\right) \\
=&\left[\frac{1}{T-t'}\right]^2\sum_{t>t'} E\left(E\left[\widehat{m}_1'(Z_{it,t'}) - m_1'(Z_{it,t'}) - h_n^p A_{t',[d]}|Z_{i1,t'},\ldots,Z_{iT,t'}\right]\left[m_1'(Z_{it,t'}) - E\left(m_1'(Z_{it,t'})\right)\right]\right) \\
=&\left[\frac{1}{T-t'}\right]^2\sum_{t>t'} E\left(E\left[\widehat{m}_1'(Z_{it,t'}) - m_1'(Z_{it,t'})\right]\cdot\left[m_1'(Z_{it,t'}) - E\left(m_1'(Z_{it,t'})\right)\right]\right) \to 0,
\end{aligned} \tag{A.11}$$

where the convergence to zero follows from Masry (1996, Theorem 6). Consequently,

$$\begin{aligned}
&\sqrt{n}\left(\widehat{\delta}_{t'} - h_n^p A_{t',[d]} - E[m_1'(Z_{it,t'})]\right) \\
=&\sqrt{n}(\widehat{\delta}_{t'} - \tilde{\delta}_{t'} - h_n^p A_{1,t'}) + \sqrt{n}\left[\tilde{\delta}_{t'} - E\left(m_1'(Z_{it,t'})\right)\right] \\
\to& N(0, \Phi_{t'} + \Omega_{t'}).
\end{aligned}$$

□

Proof of Theorem 4. From the equation (A.9) in Lemma 3 (restricted to a fixed time period), it follows that ($V_{it,t'} = \Delta Y_{it,t'} - m(Z_{it,t'})$)

$$
\begin{aligned}
& Cov(\sqrt{n}J^2_{t',r}(Z_{ik,t'}), \sqrt{n}J^2_{t',m}(Z_{il,t'})) \\
=& nE\left[\left(\frac{1}{n}\sum_{i=1}^{n} V_{ik,t'}(G(Z_{ik,t'}))_{r,1}\right)\left(\frac{1}{n}\sum_{i=1}^{n} V_{il,t'}(G(Z_{il,t'}))_{m,1}\right)\right] \\
=& \frac{1}{n}\sum_{i=1}^{n} E\left[\sigma_{kl,t'}\left(Z_{ik,t'}, Z_{il,t'}\right)\cdot (G(Z_{ik,t'}))_{r,1}\cdot (G(Z_{il,t'}))_{m,1}\right] \\
=& E\left[\sigma_{kl,t'}\left(Z_{ik,t'}, Z_{il,t'}\right)\cdot (G(Z_{ik,t'}))_{r,1}\cdot (G(Z_{il,t'}))_{m,1}\right] \\
=& (\widetilde{\Phi}^{(k,l)}_{t'})_{r-1,m-1} + o_p(1),
\end{aligned}
$$

where $r, m = 2, \ldots, d+1$, and $t' < k, l \leq T$. Combined with equation (A.10) and using the notation from the proof of Theorem (3), we have

$$
\begin{aligned}
& Cov\left(\sqrt{n}\left[\widehat{\delta}_{k,t'} - \tilde{\delta}_{k,t'} - h^p_n A_{t',[d]}(Z_{ik,t'})\right], \sqrt{n}\left[\widehat{\delta}_{l,t'} - \tilde{\delta}_{l,t'} - h^p_n A_{t',[d]}(Z_{il,t'})\right]\right) \\
=& Cov\left(-\sqrt{n}J^2_{t',[d]}(Z_{ik,t'}), -\sqrt{n}J^2_{t',[d]}(Z_{il,t'})\right) + o(1) \\
=& E\left[\sigma_{kl,t'}\left(Z_{ik,t'}, Z_{il,t'}\right)\cdot (G(Z_{ik,t'}))_{[d],1}\cdot [(G(Z_{il,t'}))_{[d],1}]'\right] + o_p(1) \\
=& \widetilde{\Phi}^{(k,l)}_{t'} + o_p(1).
\end{aligned}
$$

where the (r, m)-th element of $\widetilde{\Phi}^{(k,l)}_{t'}$ is $(\widetilde{\Phi}^{(k,l)}_{t'})_{r-1,m-1}$. Therefore by the same argument as in equation (A.11), it holds that

$$
\begin{aligned}
& Cov(\sqrt{n}\widehat{\delta}_{k,t'}, \sqrt{n}\widehat{\delta}_{l,t'}) \\
=& nCov\left(\left[\widehat{\delta}_{k,t'} - h^p_n A_{t',[d]}(Z_{ik,t'}) - E\left(m'_1(Z_{it,t'})\right)\right], \left[\widehat{\delta}_{l,t'} - h^p_n A_{t',[d]}(Z_{il,t'}) - E\left(m'_1(Z_{it,t'})\right)\right]\right) \\
=& nCov\left(\left[\widehat{\delta}_{k,t'} - \tilde{\delta}_{k,t'} - h^p_n A_{t',[d]}(Z_{ik,t'})\right], \left[\widehat{\delta}_{l,t'} - \tilde{\delta}_{l,t'} - h^p_n A_{t',[d]}(Z_{il,t'})\right]\right) \\
& + nCov\left(\left[\widehat{\delta}_{k,t'} - \tilde{\delta}_{k,t'} - h^p_n A_{t',[d]}(Z_{ik,t'})\right], \left[\tilde{\delta}_{l,t'} - E\left(m'_1(Z_{it,t'})\right)\right]\right) \\
& + nCov\left(\left[\tilde{\delta}_{k,t'} - E\left(m'_1(Z_{it,t'})\right)\right], \left[\widehat{\delta}_{l,t'} - \tilde{\delta}_{l,t'} - h^p A_{t',[d]}(Z_{il,t'})\right]\right) \\
& + nCov\left(\left[\tilde{\delta}_{k,t'} - E\left(m'_1(Z_{it,t'})\right)\right], \left[\tilde{\delta}_{l,t'} - E\left(m'_1(Z_{it,t'})\right)\right]\right) \\
=& \widetilde{\Phi}^{(k,l)}_{t'} + nCov\left(\left[\tilde{\delta}_{k,t'} - E\left(m'_1(Z_{it,t'})\right)\right], \left[\tilde{\delta}_{l,t'} - E\left(m'_1(Z_{it,t'})\right)\right]\right) + o_p(1) \\
=& \widetilde{\Phi}^{(k,l)}_{t'} + \frac{1}{n}E\left\{\left[\sum_{i=1}^{n}\left(m'_1(Z_{ik,t'})\right) - E\left(m'_1(Z_{it,t'})\right)\right]\left[\sum_{i=1}^{n}\left(m'_1(Z_{il,t'})\right) - E\left(m'_1(Z_{it,t'})\right)\right]'\right\} + o_p(1) \\
=& \widetilde{\Phi}^{(k,l)}_{t'} + \widetilde{\Omega}^{(k,l)}_{t'},
\end{aligned}
$$

where $\widetilde{\Omega}_{t'}^{(k,l)} = Cov\left(m_1'(Z_{ik,t'}), m_1'(Z_{il,t'})\right)$.

According to Theorem 3, $\sqrt{n}\overline{\delta}_{k,t'} \to N\left(0, \widetilde{\Phi}_{t'}^{(k,k)} + \widetilde{\Omega}_{t'}^{(k,k)}\right)$, where $\widetilde{\Omega}_{t'}^{(k,k)} = Var\left[m_1'(Z_{it,t'})\right]$ and $\widetilde{\Phi}_{t'}^{(k,k)} = E\left[\sigma_{t'}^2(Z_{it,t'})(G(Z_{it,t'}))_{[d],1}\left[(G^s(Z_{it,t'}))_{[d],1}\right]'\right]$. Using the Crammer-Wold device, it directly follows that $\sqrt{n}\overline{\delta}_{t'}^* \to N(0, \widetilde{\Phi}_{t'} + \widetilde{\Omega}_{t'})$. □

References

Abrevaya, J., 1999. Rank estimation of a transformation model with observed truncation. The Econometrics Journal 2 (2), 292–305.

Bester, C., Hansen, C., 2009a. Identification of marginal effects in a nonparametric correlated random effects model. Journal of Business and Economic Statistics 27 (2), 235–250.

Bester, C., Hansen, C., 2009b. A penalty function approach to bias reduction in nonlinear panel models with fixed effects. Journal of Business and Economic Statistics 27 (2), 131–148.

Blum, J., Hanson, D., Koopmans, L., 1963. On the strong law of large numbers for a class of stochastic processes. Probability Theory and Related Fields 2 (1), 1–11.

Charlier, E., Melenberg, B., van Soest, A. H. O., 1995. A smoothed maximum score estimator for the binary choice panel data model with an application to labour force participation. Statistica Neerlandica 49 (3), 324–342.

Chernozhukov, V., Fernández-Val, I., Hahn, J., Newey, W., 2013. Average and quantile effects in nonseparable panel models. Econometrica 81 (2), 535–580.

Fan, J., Gijbels, I., 1992. Variable bandwidth and local linear regression smoothers. The Annals of Statistics, 2008–2036.

Gasser, T., Müller, H., 1984. Estimating regression functions and their derivatives by the kernel method. Scandinavian Journal of Statistics, 171–185.

Hahn, J., Newey, W., 2004. Jackknife and analytical bias reduction for nonlinear panel models. Econometrica 72 (4), 1295–1319.

Härdle, W., Stoker, T., 1989. Investigating smooth multiple regression by the method of average derivatives. Journal of the American Statistical Association, 986–995.

Hoderlein, S., White, H., 2012. Nonparametric identification in nonseparable panel data models with generalized fixed effects. Journal of Econometrics 168 (2), 300–314.

Honoré, B., 1992. Trimmed lad and least squares estimation of truncated and censored regression models with fixed effects. Econometrica: Journal of the Econometric Society, 533–565.

Hristache, M., Juditsky, A., Spokoiny, V., 2001. Direct estimation of the index coefficient in a single-index model. The Annals of Statistics 29 (3), 595–623.

Ichimura, H., 1993. Semiparametric least squares (sls) and weighted sls estimation of single-index models. Journal of Econometrics 58 (1), 71–120.

Kyriazidou, E., 1997. Estimation of a panel data sample selection model. Econometrica: Journal of the Econometric Society, 1335–1364.

Lancaster, T., 2000. The incidental parameter problem since 1948. Journal of econometrics 95 (2), 391–413.

Li, Q., Lu, X., Ullah, A., 2003. Multivariate local polynomial regression for estimating average derivatives. Journal of Nonparametric Statistics 15 (4-5), 607–624.

Li, Q., Racine, J., 2007. Nonparametric econometrics: theory and practice. Princeton University Press, New Jersey.

Manski, C. F., 1987. Semiparametric analysis of random effects linear models from binary panel data. Econometrica: Journal of the Econometric Society, 357–362.

Masry, E., 1996. Multivariate local polynomial regression for time series: uniform strong consistency and rates. Journal of Time Series Analysis 17 (6), 571–599.

Masry, E., 1997. Local polynomial estimation of regression functions for mixing processes. Scandinavian Journal of Statistics 24 (2), 165–179.

Newey, W., Stoker, T., 1993. Efficiency of weighted average derivative estimators and index models. Econometrica: Journal of the Econometric Society, 1199–1223.